

# TIME ASYMPTOTIC HIGH ORDER SCHEMES FOR DISSIPATIVE BGK HYPERBOLIC SYSTEMS

DENISE AREGBA-DRIOLLET<sup>1</sup>, MAYA BRIANI<sup>2</sup>, AND ROBERTO NATALINI<sup>2</sup>

**ABSTRACT.** We introduce a new class of finite differences schemes to approximate one dimensional dissipative semilinear hyperbolic systems with a BGK structure. Using precise analytical time-decay estimates of the local truncation error, it is possible to design schemes, based on the standard upwind approximation, which are increasingly accurate for large times when approximating small perturbations of constant asymptotic states. Numerical tests show their better performances with respect to those of other schemes.

## 1. INTRODUCTION

Consider a BGK system in one space dimension, for the unknowns  $f^i \in \mathbb{R}^k$ ,  $k \geq 1$  and  $i = 1, \dots, m$ :

$$(1) \quad \begin{cases} \partial_t f^i + \lambda_i \partial_x f^i = M_i(u) - f^i, \\ \text{where } u := \sum_{i=1}^m f^i. \end{cases}$$

Here  $x \in \mathbb{R}$  and  $t > 0$ , the  $\lambda_i$ , for  $i = 1, \dots, m$ , are given distinct real values, and the functions  $M_i = M_i(u) \in \mathbb{R}^k$  are smooth functions of  $u$  such that:

$$\sum_{i=1}^m M_i(u) = u.$$

Following [3], we rewrite system (1) in its *conservative-dissipative* form. This means that we assume that there exists an invertible matrix

$$(2) \quad D = \begin{pmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{pmatrix},$$

such that, setting  $m_1 = k$ ,  $m_2 = k(m-1)$ , the new unknown

$$(3) \quad Z = Df = (u, \tilde{Z})^T \in \mathbb{R}^{m_1} \times \mathbb{R}^{m_2},$$

solves the system

$$(4) \quad \begin{cases} \partial_t u + A_{11} \partial_x u + A_{12} \partial_x \tilde{Z} = 0, \\ \partial_t \tilde{Z} + A_{21} \partial_x u + A_{22} \partial_x \tilde{Z} = \tilde{Q}(u) - \tilde{Z}, \end{cases}$$

where  $A$  is symmetric and  $\tilde{Q}(u)$  is quadratic in  $u$ , i.e.:  $\tilde{Q}(0) = 0$  and  $\tilde{Q}'(0) = 0$ . Observe that, after the transformation, the source term is zero in the first component and the second one is the sum of a quadratic term and of the dissipative term  $-\tilde{Z}$ .

It is proved in [13] and [3] that, under some additional conditions, usually induced by suitable entropy functions, and for initial data which are small perturbations of

---

2010 *Mathematics Subject Classification.* Primary: 65M12; Secondary: 35L65.

*Key words and phrases.* Finite differences methods, dissipative hyperbolic problems, BGK systems, asymptotic behavior, asymptotic high order schemes.

<sup>1</sup>IMB, UMR CNRS 5251, Université de Bordeaux.

<sup>2</sup> Istituto per le Applicazioni del Calcolo “Mauro Picone”, Consiglio Nazionale delle Ricerche.

constant equilibrium states and smooth in some suitable norms, the corresponding smooth solutions exist globally and their  $L^\infty$ -norm decay, for large times, as

$$(5) \quad u = O(t^{-1/2}), \quad \tilde{Z} = O(t^{-1}),$$

and similar estimates are available for their space and time derivatives. Notice that the improved estimate for the unknown  $\tilde{Z}$  can only be obtained in these new coordinates and does not hold for other combinations of the unknowns.

The aim of this paper is to take advantage by these time decay estimates to build up more accurate numerical schemes. To be more specific, we show in the following that for standard numerical schemes, for instance upwind schemes with the source term approximated pointwise by the standard Euler scheme, the local truncation error for the conservative-dissipative unknowns  $(u, \tilde{Z})$  has the following decay as  $t \rightarrow +\infty$ , for a fixed CFL ratio:

$$\mathcal{T}_u(x, t) = O(\Delta x \, t^{-3/2}), \quad \mathcal{T}_{\tilde{Z}}(x, t) = O(\Delta x \, t^{-3/2}).$$

It can be seen numerically that the corresponding absolute errors, for a fixed space step, decays as

$$e_u(t) = O(t^{-1/2}), \quad e_{\tilde{Z}}(t) = O(t^{-1}),$$

which implies, taking into account (5), that the relative error is essentially constant in time.

Here, our main goal is to improve the decay rate of the truncation error to achieve an effective decay in time of the relative error, both in  $u$  and  $\tilde{Z}$ . Before presenting our strategy and our main results, let us review some different attempts to design effective numerical approximations for hyperbolic equations with a source term. Let us mention some families of schemes, sometimes overlapping: Well Balanced [12, 7, 8, 9, 11, 15, 5], Runge-Kutta IMEX [20], upwinding source [21, 2, 6, 1], and asymptotic preserving [14, 16]. The main idea in all these schemes is to use some knowledge of the actual time behavior of the solutions to improve their numerical approximation, at least in some specific regimes.

In particular, in [1], the linear version of the present problem was considered and therefore, to approximate the solutions around non constant asymptotic states, some schemes were proposed, which had the property to become higher order (in space) for large times, thanks to the careful consideration of the analytical decay rates of the solutions. A different attempt was given by the well balanced schemes, see for instance [12, 7, 5], namely schemes which are exact when computed on stationary solutions of the problem, even if up to now, the time decay rate of the unsteady solutions has not yet been explicitly considered. However, the Asymptotic Preserving properties of some Well Balanced schemes, can yield nice results for large times, as in [11], see Section 7 below.

In the present case, a striking difference with these previous works lies in the fact that the asymptotic equilibrium states are constant and therefore all the consistent schemes are *exact* on them. So, the goal of our work is a bit different. In this paper, we design schemes which are able to improve their performance for large times, when the initial data are small perturbations of a given constant equilibrium state. To obtain these results, we use the estimates in [3] to perform a detailed analysis of the behavior of the truncation error for a general class of schemes, which generalize and improve those introduced in [1]. Thanks to this analysis, we are able to construct schemes such that the truncation order behaves as

$$\mathcal{T}_u(x, t) = O(\Delta x \, t^{-2}), \quad \mathcal{T}_{\tilde{Z}}(x, t) = O(\Delta x \, t^{-2}),$$

for a fixed CFL ratio and such that their asymptotic numerical error, observed in the practical tests, improves of  $t^{-1/2}$  on the previous schemes.

The plan of the paper is the following. In Section 2, we introduce our analytical framework. The main schemes are derived in Section 3, where we show how to improve the time decay of their local truncation error. Sections 4 and 5 are devoted to the monotonicity conditions for the new scheme in the  $2 \times 2$  and  $3 \times 3$  cases respectively. Then we present some remarks in the linear  $2 \times 2$  case, to allow a direct comparison with other schemes. Section 7 presents some numerical tests which show the nice behavior of our new schemes both in the linear and the nonlinear cases.

## 2. THE ANALYTICAL FRAMEWORK

Let us observe that when transforming system (1) in system (4), we can always assume that the block  $D_{11}$  and  $D_{12}$  have the special form

$$D_{11} = I_k, \quad D_{12} = (I_k I_k \cdots I_k) \in \mathbb{R}^{k \times m_2},$$

and setting  $\Lambda = \text{diag}(\lambda_1 I_k, \dots, \lambda_m I_k)$ , we have that

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} = D \Lambda D^{-1}.$$

Therefore, we can rewrite our system in a more compact form:

$$(6) \quad \partial_t Z + A \partial_x Z = -Z + DM(u).$$

To guarantee the existence of the matrix  $D$  in (2), we can assume that our system is strictly entropy dissipative in the sense of [13] and verifies the Shizuta-Kawashima condition [22, 13, 3]. For instance, following Bouchut [4], we may assume the following sufficient condition.

**Condition ED (Entropy Dissipation condition)** *There exists an open set  $\Omega \subseteq \mathbb{R}^k$  and a strictly convex function  $\eta = \eta(u) : \Omega \rightarrow \mathbb{R}$ , such that the matrix  $M'_i(u)^T \eta''(u)$  is symmetric and strictly positive defined for all  $u \in \Omega$  and  $i = 1, \dots, m$ .*

Under this condition, and using the results in [4, 13, 3], it is possible to prove the existence of a matrix  $D$  in (2), with all the properties mentioned above. The details of the derivation of the actual coefficients of this matrix, which in general is not unique, but depends on the entropy dissipative function, can be found in the general case in [3], and are not relevant for the following discussion, even if of course they are relevant to derive the final schemes. However, in the examples presented below, the matrix  $D$  is always explicitly given.

**2.1. Algebraic conditions.** Let us state now some general properties for the matrix  $D$  that we shall use in the following, and which are easily derived under **condition ED**.

First of all, since  $A = D \Lambda D^{-1}$  in (6) is symmetric, we have that the matrix  $H := D^T D$  and the diagonal matrix  $\Lambda$  commute, i.e.

$$(7) \quad D^T D \Lambda = \Lambda D^T D,$$

So, for the generic  $ij$  block element of the matrix  $H$ , we have that  $(\lambda_i - \lambda_j)H_{ij} = 0$ , hence

$$H_{ij} = 0 \text{ for } i \neq j \text{ once } \lambda_i \neq \lambda_j.$$

Since  $\Lambda$  is a block diagonal matrix with  $m$  distinct eigenvalues with multiplicity equal to  $k$ , we get that  $H$  is a block diagonal matrix of the following form

$$(8) \quad H = D^T D = \text{diag}(h_1, \dots, h_m), \quad h_i \in \mathbb{R}^{k \times k}, \quad i = 1, \dots, m.$$

Moreover, the matrix  $H = D^T D$  is symmetric and invertible and we can find the following expression for  $D^{-1}$ ,

$$(9) \quad D^{-1} = H^{-1} D^T = \begin{pmatrix} H_1^{-1} & H_1^{-1} D_{21}^T \\ H_2^{-1} D_{12}^T & H_2^{-1} D_{22}^T \end{pmatrix},$$

where  $H_1 = h_1$  and  $H_2 = \text{diag}(h_2, \dots, h_m)$ . Finally, using (8), the following relations hold:

$$(10) \quad D_{12} = -D_{21}^T D_{22}, \quad D_{21}^T D_{21} = H_1 - I_k, \quad D_{12}^T D_{12} + D_{22}^T D_{22} = H_2.$$

There is another important relation involving the matrix  $H$ , which will be of use in the following. Let  $A_{11}$  be the top-left block of the symmetric matrix  $A$  in system (6). Therefore:  $A_{11} = \Lambda_1 H_1^{-1} + D_{12} \Lambda_2 H_2^{-1} D_{12}^T$ , i.e.

$$(11) \quad A_{11} = \lambda_1 h_1^{-1} + \left( I_k I_k \cdots I_k \right) \text{diag}(\lambda_2 h_2^{-1}, \dots, \lambda_m h_m^{-1}) \begin{pmatrix} I_k \\ I_k \\ \vdots \\ I_k \end{pmatrix} = \sum_{i=1}^m \lambda_i h_i^{-1}.$$

From now on, we shall consider each matrix  $N \in \mathbb{R}^{km \times km}$  to be decomposed into blocks as follows:

$$(12) \quad N = \begin{pmatrix} N_{11} & N_{12} \\ N_{21} & N_{22} \end{pmatrix},$$

with  $N_{11} \in \mathbb{R}^{k \times k}$ ,  $N_{12} \in \mathbb{R}^{k \times m_2}$ ,  $N_{21} \in \mathbb{R}^{m_2 \times k}$ ,  $N_{22} \in \mathbb{R}^{m_2 \times m_2}$ . For the particular case of a block diagonal matrix  $N = \text{diag}(n_1, \dots, n_m)$ ,  $n_i \in \mathbb{R}^{k \times k}$  for  $i = 1, \dots, m$ , we shall use the following notation

$$(13) \quad N = \text{diag}(N_1, N_2), \quad N_1 = n_1 \text{ and } N_2 = \text{diag}(n_2, \dots, n_m).$$

Moreover, for a generic vector  $V \in \mathbb{R}^{km}$  we shall write

$$(14) \quad V = (V_1, \tilde{V}), \quad \text{with } V_1 \in \mathbb{R}^k, \quad \tilde{V} \in \mathbb{R}^{m_2}.$$

**Example 2.1.** Consider the special case of the  $2 \times 2$  hyperbolic Jin-Xin relaxation system [17, 19]:

$$(15) \quad \begin{cases} \partial_t u + \partial_x v = 0, \\ \partial_t v + \lambda^2 \partial_x u = F(u) - v, \end{cases}$$

for  $\lambda > 0$ , where the unknowns  $u$  and  $v$  are scalar and the function  $F = F(u)$  is smooth, with  $F(0) = 0$ . This case is obtained from (1) for  $k = 1$ ,  $m = 2$ , and  $\lambda_2 = -\lambda_1 = \lambda$ , by setting

$$u = f^1 + f^2, \quad v = \lambda(f^2 - f^1), \quad F(u) = \lambda(M_2 - M_1).$$

Recall that for bounded initial data, and for  $\lambda > \tilde{M}$ , where  $\tilde{M}$  is a positive constant which depends on  $F$  and on the initial data, there exists a global bounded solution to the Cauchy problem (15), see [19]. Under the weaker condition

$$(16) \quad \lambda > |F'(0)|,$$

the problem is dissipative in the sense of [13] and the Shizuta-Kawashima condition is verified, at least for small values of  $u$ , and so, at least for smooth and small initial data, there exists again a global smooth solution to the same problem. To obtain the time decay rates of these solutions, we need to rewrite the problem in the *conservative-dissipative* coordinates. Assume (16) and set  $a = F'(0)$  so that the

quantity  $\mu = (\lambda^2 - a^2)^{-1/2}$  is real and positive. Setting  $\tilde{Z} = \mu(v - au)$ , the new unknown  $(u, \tilde{Z})$  solves the problem in form (4):

$$(17) \quad \begin{cases} \partial_t u + \partial_x(au + \frac{1}{\mu}\tilde{Z}) = 0, \\ \partial_t \tilde{Z} + \partial_x(\frac{u}{\mu} - a\tilde{Z}) = \mu(F(u) - au) - \tilde{Z}. \end{cases}$$

In this case the matrix  $D$  is given by

$$D = \begin{pmatrix} 1 & 1 \\ -\mu a_+ & \mu a_- \end{pmatrix},$$

where  $a_{\pm} = \lambda \pm a > 0$ , from assumption (16).

**Example 2.2.** Let us now compute the conservative-dissipative form for the following  $3 \times 3$  BGK model,

$$(18) \quad \begin{cases} \partial_t f_1 - \lambda \partial_x f_1 = M_1(u) - f_1, \\ \partial_t f_2 = M_2(u) - f_2, \\ \partial_t f_3 + \lambda \partial_x f_3 = M_3(u) - f_3. \end{cases}$$

Let  $F = F(u)$  be a smooth scalar function such that  $F(0) = 0$  and let  $\gamma$  be such that  $\gamma'(u) = |F'(u)|$ , with  $\gamma(0) = 0$ . We choose our three maxwellian functions as follows, for  $\beta \in ]0, 1[$  and  $\lambda > 0$

$$(19) \quad M_1(u) = \frac{1}{2} \left( \frac{\gamma(u) - F(u)}{\lambda} + \beta u \right), \quad M_3(u) = \frac{1}{2} \left( \frac{\gamma(u) + F(u)}{\lambda} + \beta u \right),$$

$$M_2(u) = u - M_1(u) - M_3(u) = (1 - \beta)u - \frac{\gamma(u)}{\lambda}.$$

The functions  $M_i$ ,  $i = 1, 2, 3$ , are strictly increasing if for any  $u$  under consideration

$$(20) \quad \lambda > \frac{|F'(u)|}{1 - \beta},$$

and so **condition ED** is verified. Let  $a = F'(0)$  and  $\alpha = |a| + \beta\lambda$ , following the results in [4, 13, 3] it is possible to compute the matrix  $D$  for the change of variables (3) as

$$D = \begin{pmatrix} 1 & 1 & 1 \\ \frac{\alpha+a}{\alpha-a} \sqrt{\frac{\lambda(\alpha-a)}{\alpha(\alpha+a)}} & 0 & -\sqrt{\frac{\lambda(\alpha-a)}{\alpha(\alpha+a)}} \\ -\sqrt{\frac{\lambda-\alpha}{\alpha}} & -\sqrt{\frac{\lambda-\alpha}{\alpha}} + \frac{\lambda}{\sqrt{\alpha(\lambda-\alpha)}} & -\sqrt{\frac{\lambda-\alpha}{\alpha}} \end{pmatrix}.$$

**2.2. Time Decay Properties.** Here we report some time decay results which have been mainly proved in [3], and which will be useful in the following.

**Theorem 2.3.** *Let  $Z = (u, \tilde{Z})$  be the local smooth solution to the Cauchy problem for system (4). Let  $E_s = \max \{\|Z(0)\|_{L^1}, \|Z(0)\|_{H^s}\}$  a norm for the initial data, which are taken in  $H^k$  for  $k$  large enough, and assume  $E_2$  small enough. Therefore, under the Condition ED, the solution is global in time and the following decay estimate holds, for all  $\beta \leq k$ :*

$$(21) \quad \|\partial_x^\beta Z(t)\|_{L^\infty} \leq C \min \{1, t^{-\frac{1}{2} - \frac{\beta}{2}}\} E_{\beta+1},$$

with  $C = C(E_{\beta+\sigma})$  for  $\sigma$  large enough. For the dissipative part  $\tilde{Z}$  we have, under the same conditions, the more precise estimate

$$(22) \quad \|\partial_x^\beta \tilde{Z}(t)\|_{L^\infty} \leq C \min \{1, t^{-1 - \frac{\beta}{2}}\} E_{\beta+1},$$

for another constant  $C = C(E_{\beta+\sigma})$  as previously.

**Theorem 2.4.** *Under the assumptions of Theorem 2.3, we have the following decay estimates for the time derivatives of  $Z$ :*

$$(23) \quad \|\partial_x^\beta Z_t(t)\|_{L^\infty} \leq C \min\{1, t^{-1-\frac{\beta}{2}}\} E_{\beta+2};$$

$$(24) \quad \|\partial_x^\beta \tilde{Z}_t(t)\|_{L^\infty} \leq C \min\{1, t^{-\frac{3}{2}-\frac{\beta}{2}}\} E_{\beta+3};$$

$$(25) \quad \|\partial_x^\beta Z_{tt}(t)\|_{L^\infty} \leq C \min\{1, t^{-\frac{3}{2}-\frac{\beta}{2}}\} E_{\beta+3};$$

with  $C = C(E_{\beta+\sigma})$  for  $\sigma$  large enough.

*Proof of Theorem (2.4).* We have only to prove inequality (25), since the other ones are proved in [3]. First we have that

$$(26) \quad \partial_{tt}u = A_{11}^2 \partial_{xx}u + A_{11}A_{22} \partial_{xx}\tilde{Z} - A_{12} \partial_{tx}\tilde{Z},$$

and so, using inequalities (22), (23) and (24), we have (25) for the conservative part. Now, using the second part of system (4), we have

$$(27) \quad \partial_{tt}\tilde{Z} = -A_{21} \partial_{tx}u - A_{22} \partial_{tx}\tilde{Z} + \tilde{Q}'(u) \partial_t u - \partial_t \tilde{Z},$$

which yields the proof, using that the term  $\tilde{Q}(u)$  is at least quadratic in  $u$ .  $\square$

**Remark 2.5.** Consider the special case of system (4) when  $A_{11} = 0$ . We can improve the decay of the time derivative of the unknowns. Actually

$$\|u_t(t)\|_{L^\infty} = \|A_{12} \partial_x \tilde{Z}\|_{L^\infty} \leq C \min\{1, t^{-\frac{3}{2}}\} E_3.$$

Then, we have

$$(28) \quad \|\partial_{tt}u(t)\|_{L^\infty} = \|A_{12} \partial_{tx}\tilde{Z}\|_{L^\infty} \leq C \min\{1, t^{-2}\} E_4,$$

$$(29) \quad \|\partial_{tx}u(t)\|_{L^\infty} = \|A_{12} \partial_{xx}\tilde{Z}\|_{L^\infty} \leq C \min\{1, t^{-2}\} E_4.$$

Finally, for the second derivative of the dissipative part, we have

$$(30) \quad \partial_{tt}\tilde{Z} = -A_{21} \partial_{tx}u - A_{22} \partial_{tx}\tilde{Z} + \tilde{Q}'(u) \partial_t u - \partial_t \tilde{Z}.$$

Therefore, we can apply the Duhamel's formula to obtain

$$\partial_t \tilde{Z} = e^{-t} \partial_t \tilde{Z}_0 - \int_0^t e^{-(t-s)} \left( A_{21} \partial_{tx}u(s) + A_{22} \partial_{tx}\tilde{Z}(s) + \tilde{Q}'(u) \partial_t u(s) \right) ds.$$

Now, since the function  $Q$  is quadratic in  $u$  and using the previous estimates, we find that

$$\|\partial_t \tilde{Z}\|_{L^\infty} \leq C \left( e^{-t} + \int_0^t e^{-(t-s)} \min\{1, s^{-2}\} ds \right) \leq C \min\{1, t^{-2}\}.$$

By plugging this last inequality in (30) we obtain

$$(31) \quad \|\partial_{tt}\tilde{Z}\|_{L^\infty} + \|\partial_t \tilde{Z}\|_{L^\infty} \leq C \min\{1, t^{-2}\}.$$

### 3. THE NUMERICAL APPROXIMATION

In this section we first introduce general finite difference approximations for system (1). Then, we compute the local truncation error of these schemes and we discuss its decays properties. The main result is given in Theorem 3.1, where a class of TAHO schemes is fully characterized. First, we approximate the differential part following the direction of the characteristic velocities, so we study the methods for the system in diagonal form (1).

We denote by  $f = (f^1, \dots, f^m)$  the exact solution. Let  $\Delta x$  the uniform mesh-length and  $x_j = j \Delta x$  the spatial grid points for all  $j \in \mathbb{Z}$ . The time levels  $t_n$ , with  $t_0 = 0$ , are also spaced uniformly with mesh-length  $\Delta t = t_{n+1} - t_n$  for  $n \in \mathbb{N}$ . We

denote by  $\rho$  the CFL ratio  $\rho = \Delta t / \Delta x$ , which is taken constant through all the paper.

We consider the Cauchy problem for system (1) possibly subjected to some stability conditions. The initial data  $f^0$  is supposed to be smooth and approximated by its node values. The approximate solution  $(f_{j,n}^1, \dots, f_{j,n}^m)^T$ ,  $f_{j,n}^i \in \mathbb{R}^{m_1}$ ,  $i = 1, \dots, m$ , for  $j \in \mathbb{Z}$  and  $n \in \mathbb{N}$ , is given by

$$(32) \quad \begin{aligned} & \frac{f_{j,n+1}^i - f_{j,n}^i}{\Delta t} + \frac{\lambda_i}{2\Delta x} (f_{j+1,n}^i - f_{j-1,n}^i) - \frac{q_i}{2\Delta x} \delta_x^2 f_{j,n}^i \\ &= \sum_{l=-1,0,1} (\mathcal{B}_l^i(u_{j+l,n}) - \beta_l^i f_{j+l,n}^i), \end{aligned}$$

$$f_{j,0}^i = f_0^i(x_j), \quad j \in \mathbb{Z},$$

where  $\delta_x^2 f_{j,n}^i = (f_{j+1,n}^i - 2f_{j,n}^i + f_{j-1,n}^i)$ , for all  $i = 1, \dots, m$ . The artificial diffusion terms  $q_i$  are diagonal matrices in  $\mathbb{R}_+^{m_1 \times m_1}$ . The source term approximation is defined, for  $l = -1, 0, 1$ , by the diagonal matrices  $\beta_l^i \in \mathbb{R}^{m_1 \times m_1}$  and by the vectors of functions  $\mathcal{B}_l^i(\cdot) \in \mathbb{R}^{m_1}$ .

We assume the scheme (32) is consistent with system (1), i.e, for all  $i = 1, \dots, m$

$$(33) \quad \begin{aligned} & \beta_{-1}^i + \beta_0^i + \beta_1^i = I_{m_1} + \Delta x C^i, \\ & \mathcal{B}_{-1}^i(u) + \mathcal{B}_0^i(u) + \mathcal{B}_1^i(u) = M_i(u) + \Delta x \mathcal{C}_i(u), \end{aligned}$$

where  $C^i = \text{diag}(c_1^i, \dots, c_{m_1}^i) \in \mathbb{R}^{m_1 \times m_1}$  and  $\mathcal{C}_i(u)$  are  $m_1$  functions to be defined.

By applying the change of variables (3), the scheme applies to the system in the conservative-dissipative form (4), and it reads

$$(34) \quad \begin{aligned} & \frac{Z_{j,n+1} - Z_{j,n}}{\Delta t} + \frac{A}{2\Delta x} (Z_{j+1,n} - Z_{j-1,n}) - \frac{\bar{Q}}{2\Delta x} \delta_x^2 Z_{j,n} \\ &= \sum_{l=-1,0,1} (\bar{\mathcal{B}}_l(u_{j+l,n}) - \bar{b}_l Z_{j+l,n}), \end{aligned}$$

where  $Q = \text{diag}(q_1, \dots, q_m)$ ,  $\bar{Q} = DQD^{-1}$  and for  $l = -1, 0, 1$ ,

$$b_l = \text{diag}(\beta_l^{1,1}, \dots, \beta_l^{1,k}, \dots, \beta_l^{m,1}, \dots, \beta_l^{m,k}) \in \mathbb{R}^{km \times km}, \quad \bar{b}_l = Db_l D^{-1} \in \mathbb{R}^{km \times km},$$

$$\begin{aligned} \mathcal{B}_l(u) &= (\mathcal{B}_l^{1,1}(u), \dots, \mathcal{B}_l^{1,k}(u), \dots, \mathcal{B}_l^{m,1}(u), \dots, \mathcal{B}_l^{m,k}(u))^T \in \mathbb{R}^{km}, \\ \bar{\mathcal{B}}_l &= D\mathcal{B}_l(u) \in \mathbb{R}^{km}. \end{aligned}$$

By separating the system with respect to the two variables  $u$  and  $\tilde{Z}$ , we get

$$(35) \quad \begin{aligned} & \frac{u_{j,n+1} - u_{j,n}}{\Delta t} + \frac{A_{11}}{2\Delta x} (u_{j+1,n} - u_{j-1,n}) + \frac{A_{12}}{2\Delta x} (\tilde{Z}_{j+1,n} - \tilde{Z}_{j-1,n}) \\ & - \frac{\bar{Q}_{11}}{2\Delta x} \delta_x^2 u_{j,n} - \frac{\bar{Q}_{12}}{2\Delta x} \delta_x^2 \tilde{Z}_{j,n} = \sum_{l=-1,0,1} (\bar{\mathcal{B}}_l^1(u_{j+l,n}) - \bar{b}_l^{11} u_{j+l,n} - \bar{b}_l^{12} \tilde{Z}_{j+l,n}), \\ & \frac{\tilde{Z}_{j,n+1} - \tilde{Z}_{j,n}}{\Delta t} + \frac{A_{21}}{2\Delta x} (u_{j+1,n} - u_{j-1,n}) + \frac{A_{22}}{2\Delta x} (\tilde{Z}_{j+1,n} - \tilde{Z}_{j-1,n}) \\ & - \frac{\bar{Q}_{21}}{2\Delta x} \delta_x^2 u_{j,n} - \frac{\bar{Q}_{22}}{2\Delta x} \delta_x^2 \tilde{Z}_{j,n} = \sum_{l=-1,0,1} (\bar{\mathcal{B}}_l^2(u_{j+l,n}) - \bar{b}_l^{21} u_{j+l,n} - \bar{b}_l^{22} \tilde{Z}_{j+l,n}). \end{aligned}$$

where, for  $l = -1, 0, 1$ , vector  $\bar{\mathcal{B}}_l(u) = (\bar{\mathcal{B}}_l^1(u), \widetilde{\bar{\mathcal{B}}_l}(u))^T$ , follows the notation given in (14) and the block decomposition of matrices  $\bar{b}_l$  and  $\bar{Q}$  follows the definition given in (12).

**3.1. Decay properties of the local truncation error.** In this section we focus on the local truncation error for the general scheme (34). By applying the time decay properties given in Section 2.2, we will show how it is possible to build up numerical schemes which are more accurate for large times.

Set, for  $i = 1, \dots, m$ ,

$$(36) \quad \begin{aligned} C &= \text{diag}(C^i), \quad \bar{C} = DCD^{-1}, \quad \mathcal{C}(u) = (\mathcal{C}_i(u))^T, \quad \gamma^i = (\beta_1^i - \beta_{-1}^i), \\ \bar{G} &= (\bar{b}_1 - \bar{b}_{-1}), \quad \Gamma = (\mathcal{B}_1(u) - \mathcal{B}_{-1}(u)), \quad \bar{\Gamma} = (\bar{\mathcal{B}}_1(u) - \bar{\mathcal{B}}_{-1}(u)). \end{aligned}$$

For  $i = 1, \dots, m$ , we have

$$\bar{\mathcal{B}}_l(u_{j+l,n}) = \bar{\mathcal{B}}_l(u_{j,n}) + l \Delta x \bar{\mathcal{B}}'_l(u_{j,n}) \partial_x u(x_j, t_n) + O(\Delta x^2),$$

where  $\bar{\mathcal{B}}'_l \in \mathbb{R}^{km \times k}$  is the Jacobian matrix of  $\bar{\mathcal{B}}_l$ . Using the Taylor expansion and the consistency property (33), the local truncation error for the scheme (35) becomes

$$(37) \quad \begin{aligned} \mathcal{T}_u &= \frac{\Delta t}{2} \partial_{tt} u - \Delta x \frac{\bar{Q}_{11}}{2} \partial_{xx} u - \Delta x \frac{\bar{Q}_{12}}{2} \partial_{xx} \tilde{Z} \\ &\quad + \Delta x \left[ T_0^u(u) + T_1^u \partial_x u + S_0^u \tilde{Z} + S_1^u \partial_x \tilde{Z} \right] + O(\Delta t^2 + \Delta x^2), \end{aligned}$$

and

$$(38) \quad \begin{aligned} \mathcal{T}_z &= \frac{\Delta t}{2} \partial_{tt} \tilde{Z} - \Delta x \frac{\bar{Q}_{21}}{2} \partial_{xx} u - \Delta x \frac{\bar{Q}_{22}}{2} \partial_{xx} \tilde{Z} \\ &\quad + \Delta x \left[ T_0^z(u) + T_1^z \partial_x u + S_0^z \tilde{Z} + S_1^z \partial_x \tilde{Z} \right] + O(\Delta t^2 + \Delta x^2), \end{aligned}$$

where

$$(39) \quad T_0^u(u) = - \left( D_{11} \mathcal{C}^1(u) + D_{12} \tilde{\mathcal{C}}(u) - \bar{C}_{11} u \right) \in \mathbb{R}^k, \quad T_1^u = -\bar{\Gamma}'_1 + \bar{G}_{11} \in \mathbb{R}^{k \times k},$$

$$T_0^z(u) = - (D_{21} \mathcal{C}^1(u) + D_{22} \tilde{\mathcal{C}}(u) - \bar{C}_{21} u) \in \mathbb{R}^{m_2}, \quad T_1^z = (-\tilde{\Gamma}' + \bar{G}_{21}) \in \mathbb{R}^{m_2 \times k},$$

$$S_0^u = \bar{C}_{12} \in \mathbb{R}^{k \times m_2}, \quad S_1^u = \bar{G}_{12} \in \mathbb{R}^{k \times m_2},$$

$$S_0^z = \bar{C}_{22} \in \mathbb{R}^{m_2 \times m_2}, \quad S_1^z = \bar{G}_{22} \in \mathbb{R}^{m_2 \times m_2},$$

where  $\tilde{\Gamma}'$  is the  $m_2 \times k$  jacobian matrix of  $\tilde{\Gamma}$ . Clearly, the scheme (34) is at least consistent, which means it is formally of order  $O(\Delta x + \Delta t)$ . Taking  $\Gamma_i = 0$ ,  $\gamma^i = 0$ ,  $C^i = 0$  and  $\mathcal{C}_i = 0$ , for  $i = 1, \dots, m$ , we get the standard upwind scheme with the pointwise approximation for the source term and the local truncation error is just given by

$$\mathcal{T}_u(x, t) = \frac{\Delta t}{2} \partial_{tt} u - \Delta x \frac{\bar{Q}_{11}}{2} \partial_{xx} u - \Delta x \frac{\bar{Q}_{12}}{2} \partial_{xx} \tilde{Z} + O(\Delta x^2 + \Delta t^2),$$

$$\mathcal{T}_z(x, t) = \frac{\Delta t}{2} \partial_{tt} \tilde{Z} - \Delta x \frac{\bar{Q}_{21}}{2} \partial_{xx} u - \Delta x \frac{\bar{Q}_{22}}{2} \partial_{xx} \tilde{Z} + O(\Delta x^2 + \Delta t^2).$$

Using the time decay estimates in Theorems 2.3 and 2.4, we obtain for a general approximation the following estimates for the local truncation error, as  $t \rightarrow +\infty$ ,

$$\mathcal{T}_u(x, t) = O(\Delta x t^{-3/2}) + O(\Delta t t^{-3/2}), \quad \mathcal{T}_z(x, t) = O(\Delta x t^{-3/2}) + O(\Delta t t^{-3/2}).$$



Starting from the general scheme (34), we would like to improve the decay property of this local truncation error to build up more accurate numerical schemes. The main idea is to chose the free parameters of the scheme to delete the terms that decay more slowly in (37)-(38), i.e. the terms which decays as  $t^{-3/2}$ .

Let  $g_i = \text{diag}(\gamma_{(i-1)m_1+1}, \dots, \gamma_{im_1})$  for  $i = 1, \dots, m$  and  $G = \text{diag}(g_1, \dots, g_m)$ .

**Theorem 3.1** (Local Truncation Error). *Let  $\Delta t/\Delta x = \rho$  be fixed and let  $H = \text{diag}(h_1, \dots, h_m)$  be the block diagonal matrix given in (8). Recall that, by (11),  $A_{11} = \sum_{i=1}^m \lambda_i h_i^{-1}$ , and set  $P = \sum_{i=1}^m \lambda_i^2 h_i^{-1}$ . Assume  $A_{11} \neq 0$  and that the following condition holds:*

$$(40) \quad \text{the matrix } (\lambda_i I_k - A_{11}) \text{ is invertible for } i = 1, \dots, m.$$

If we make the following choice for the coefficients of the scheme (32),

$$(41) \quad C = -\frac{\rho}{2} I_{km}, \quad \mathcal{C} = CM(u) = -\frac{\rho}{2} M(u),$$

$$(42) \quad g_i = -\left(\frac{1}{2} q_i h_i^{-1} - \frac{\rho}{2} h_i^{-1} (P - (\lambda_i I_k - A_{11})^2)\right) (\lambda_i I_k - A_{11})^{-1} h_i$$

and

$$(43) \quad \Gamma'_i(u) = g_i M'_i(u) + \frac{\rho}{2} \left( (h_i^{-1} - M'_i(u)) A_{11} + \lambda_i M'_i(u) - h_i^{-1} \sum_{j=1}^m \lambda_j M'_j(u) \right),$$

both for  $i = 1, \dots, m$ , then the local truncation error of the scheme (32) decays as

$$(44) \quad \mathcal{T}_u(x, t) = O(\Delta x t^{-2}) + O(\Delta x^2 t^{-3/2}), \quad \mathcal{T}_z(x, t) = O(\Delta x t^{-2}) + O(\Delta x^2 t^{-3/2}).$$

*Proof.* Set

$$(45) \quad \begin{aligned} \tilde{T}_0 &= -\tilde{Q}(u), \\ \tilde{T}_1 &= -(A_{22}(\tilde{Q}'(u)) + (\tilde{Q}'(u))A_{11} - A_{21}), \quad \tilde{T}_2 = A_{21}A_{11} + A_{22}A_{21}, \\ \tilde{S}_1 &= 2A_{22} - (\tilde{Q}'(u))A_{12}, \quad \tilde{S}_2 = A_{21}A_{12} + A_{22}^2. \end{aligned}$$

Replacing these expressions in (37)-(38), and using the structure of the system, yields

$$(46) \quad \begin{aligned} \mathcal{T}_u &= \Delta x \left[ T_0^u(u) + (T_1^u + S_1^u(\tilde{Q}'(u))) \partial_x u + \left( \frac{\rho}{2} A_{11}^2 - \frac{\bar{Q}_{11}}{2} - S_1^u A_{21} \right) \partial_{xx} u + S_0^u \tilde{Z} \right. \\ &\quad \left. - \left( \frac{\rho}{2} A_{12} + S_1^u \right) \partial_{tx} \tilde{Z} + \left( -\frac{\bar{Q}_{12}}{2} + \frac{\rho}{2} A_{11} A_{22} - S_1^u A_{22} \right) \partial_{xx} \tilde{Z} \right] + O(\Delta x^2 t^{-3/2}), \\ \mathcal{T}_z &= \Delta x \left[ \left( \frac{\rho}{2} \tilde{T}_0(u) + T_0^z(u) \right) + \left( T_1^z + \frac{\rho}{2} \tilde{T}_1 + (S_1^z + \frac{\rho}{2} \tilde{S}_1)(\tilde{Q}'(u)) \right) \partial_x u \right. \\ &\quad \left. + \left( \frac{\rho}{2} \tilde{T}_2 - \frac{\bar{Q}_{21}}{2} - (S_1^z + \frac{\rho}{2} \tilde{S}_1) A_{21} \right) \partial_{xx} u + \left( \frac{\rho}{2} I_{m_2} + S_0^z \right) \tilde{Z} - \left( S_1^z + \frac{\rho}{2} \tilde{S}_1 \right) \partial_{tx} \tilde{Z} \right. \\ &\quad \left. + \left( \frac{\rho}{2} \tilde{S}_2 - \frac{\bar{Q}_{22}}{2} - (S_1^z + \frac{\rho}{2} \tilde{S}_1) A_{22} \right) \partial_{xx} \tilde{Z} \right] + O(\Delta x^2 t^{-3/2}), \end{aligned}$$

Our choice of the coefficients  $\mathcal{B}_l^i(u)$ ,  $\beta_l^i$ ,  $C^i$  and  $\mathcal{C}_i(u)$  is made just to cancel the coefficients of the slowly terms, i.e.

$$(47) \quad T_0^u = 0, \quad S_0^u = 0,$$

$$(48) \quad \frac{\rho}{2} \tilde{T}_0(u) + T_0^z(u) = 0, \quad \frac{\rho}{2} I_{m_2} + S_0^z = 0,$$

$$(49) \quad T_1^u + S_1^u \tilde{Q}'(u) = 0, \quad \frac{\rho}{2} A_{11}^2 - \frac{\bar{Q}_{11}}{2} - S_1^u A_{21} = 0,$$

$$(50) \quad T_1^z + \frac{\rho}{2} \tilde{T}_1 + (S_1^z + \frac{\rho}{2} \tilde{S}_1) \tilde{Q}'(u) = 0, \quad \frac{\rho}{2} \tilde{T}_2 - \frac{\bar{Q}_{21}}{2} - (S_1^z + \frac{\rho}{2} \tilde{S}_1) A_{21} = 0.$$

Therefore, the local truncation error reduces to

$$(51) \quad \begin{aligned} \mathcal{T}_u = & \Delta x \left[ - \left( \frac{\rho}{2} A_{12} + S_1^u \right) \partial_{tx} \tilde{Z} \right. \\ & \left. + \left( -\frac{\bar{Q}_{12}}{2} + \frac{\rho}{2} A_{11} A_{22} - S_1^u A_{22} \right) \partial_{xx} \tilde{Z} \right] + O(\Delta x^2 t^{-3/2}), \\ \mathcal{T}_z = & \Delta x \left[ - \left( S_1^z + \frac{\rho}{2} \tilde{S}_1 \right) \partial_{tx} \tilde{Z} \right. \\ & \left. + \left( \frac{\rho}{2} \tilde{S}_2 - \frac{\bar{Q}_{22}}{2} - (S_1^z + \frac{\rho}{2} \tilde{S}_1) A_{22} \right) \partial_{xx} \tilde{Z} \right] + O(\Delta x^2 t^{-3/2}). \end{aligned}$$

and by the estimates given in (21)-(24) the thesis is achieved.

We need now to show that system (47)-(50) has always a solution given by the relations (41)-(43). The terms given in (41) for the  $\mathcal{C}$  and  $C$  coefficients are obtained from (47) and (48). Indeed, we get

$$\begin{aligned} (-D_{11} \mathcal{C}^1(u) - D_{12} \tilde{\mathcal{C}}(u) + \bar{C}_{11} u) &= 0, \quad \bar{C}_{12} = 0, \\ (-\frac{\rho}{2} \tilde{Q}(u) + (-D_{21} \mathcal{C}^1(u) - D_{22} \tilde{\mathcal{C}}(u) + \bar{C}_{21} u)) &= 0, \quad \frac{\rho}{2} I_{m_2} + \bar{C}_{22} = 0, \end{aligned}$$

or in a more compact form,

$$- \begin{pmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{pmatrix} \begin{pmatrix} \mathcal{C}^1(u) \\ \tilde{\mathcal{C}}(u) \end{pmatrix} + \begin{pmatrix} \bar{C}_{11} & \bar{C}_{12} \\ \bar{C}_{21} & \bar{C}_{22} \end{pmatrix} \begin{pmatrix} u \\ \tilde{Z} \end{pmatrix} = \begin{pmatrix} 0 \\ \frac{\rho}{2} (\tilde{Q}(u) - I_{m_2} \tilde{Z}) \end{pmatrix},$$

which can be rewritten as

$$-D\mathcal{C} + DCD^{-1}Z = \begin{pmatrix} 0 \\ \frac{\rho}{2} (\tilde{Q}(u) - \tilde{Z}) \end{pmatrix}.$$

Now, we multiply on the left by the matrix  $D^{-1}$  to obtain

$$(52) \quad -\mathcal{C} + Cf = \frac{\rho}{2} D^{-1} \begin{pmatrix} 0 \\ \tilde{Q}(u) - \tilde{Z} \end{pmatrix} = \frac{\rho}{2} (-f + M(u)),$$

that gives (41).

We shall now focus on the derivation of the relations (42) and (43). From (49)-(50), for the matrix of free coefficients  $\bar{G}$  we have to impose

$$(53) \quad \bar{G}_{12} A_{21} = -\frac{\bar{Q}_{11}}{2} + \frac{\rho}{2} A_{11}^2, \quad \bar{G}_{22} A_{21} = -\frac{\bar{Q}_{21}}{2} + \frac{\rho}{2} (\tilde{T}_2 - \tilde{S}_1 A_{21}).$$

Now, using the notation (13) for the two diagonal matrices  $G$  and  $Q$  and by applying relations (8) and (9), we have that for  $\tilde{G} = DGD^{-1}$  and  $\tilde{Q} = DQD^{-1}$

$$\begin{aligned}\bar{G}_{12} &= G_1 H_1^{-1} D_{21}^T + D_{12} G_2 H_2^{-1} D_{22}^T, \\ \bar{G}_{22} &= D_{21} G_1 H_1^{-1} D_{21}^T + D_{22} G_2 H_2^{-1} D_{22}^T, \\ \bar{Q}_{11} &= Q_1 H_1^{-1} + D_{12} Q_2 H_2^{-1} D_{12}^T, \\ \bar{Q}_{21} &= D_{21} Q_1 H_1^{-1} + D_{22} Q_2 H_2^{-1} D_{12}^T.\end{aligned}$$

Since  $D_{11} = I_k$ , we have that (53) becomes

$$(54) \quad \begin{pmatrix} G_1 H_1^{-1} D_{21}^T A_{21} \\ G_2 H_2^{-1} D_{22}^T A_{21} \end{pmatrix} = -\frac{1}{2} \begin{pmatrix} Q_1 H_1^{-1} \\ Q_2 H_2^{-1} D_{12}^T \end{pmatrix} + \frac{\rho}{2} D^{-1} \begin{pmatrix} A_{11}^2 \\ A_{21} A_{11} - A_{22} A_{21} \end{pmatrix},$$

where we used relations (45) to sort out the term  $\tilde{T}_2 - \tilde{S}_1 A_{21} = A_{21} A_{11} - A_{22} A_{21}$ . Here we used the fact that, neglecting the terms with a faster decay, we can replace  $\tilde{S}_1$  with  $2A_{22}$ . Actually, since the term  $\tilde{S}_1$  in (46) multiplies only terms that decay faster than  $t^{-3/2}$ , it is possible to write

$$\tilde{S}_1 = 2A_{22} - \tilde{Q}'(u)A_{12} = 2A_{22} + O(\Delta x t^{-2}).$$

From (54) we get

$$(55) \quad \begin{aligned}G_1 H_1^{-1} D_{21}^T A_{21} &= -\frac{1}{2} Q_1 H_1^{-1} + \frac{\rho}{2} H_1^{-1} (A_{11}^2 + D_{21}^T (A_{21} A_{11} - A_{22} A_{21})) \\ G_2 H_2^{-1} D_{22}^T A_{21} &= -\frac{1}{2} Q_2 H_2^{-1} D_{12}^T + \frac{\rho}{2} H_2^{-1} (D_{12}^T A_{11}^2 + D_{22}^T (A_{21} A_{11} - A_{22} A_{21})),\end{aligned}$$

Using the specific form of  $D$  and relations (10), by algebraic considerations we get

$$A_{11}^2 + D_{21}^T (A_{21} A_{11} - A_{22} A_{21}) = P - (\lambda_1 I_k - A_{11})^2,$$

and

$$D_{12}^T A_{11}^2 + D_{22}^T (A_{21} A_{11} - A_{22} A_{21}) = D_{12}^T P - D_{12}^T A_{11}^2 - \Lambda_2^2 D_{12}^T + 2\Lambda_2 D_{12}^T A_{11}.$$

Then, we get for  $i = 1, \dots, m$

$$(56) \quad g_i h_i^{-1} (\lambda_i I_k - A_{11}) = -\frac{1}{2} q_i h_i^{-1} + \frac{\rho}{2} h_i^{-1} (P - (\lambda_i I_k - A_{11})^2).$$

Assuming for  $i = 1, \dots, m$ ,  $(\lambda_i I_k - A_{11})$  to be invertible, we obtain relations (42).

Finally, we need to compute the vector function  $\Gamma(u)$ . From (50) we obtain the two following relations

$$(57) \quad -\bar{\Gamma}'_1 + \bar{G}_{11} + \bar{G}_{12} \tilde{Q}'(u) = 0, \quad -\tilde{\Gamma}' + \bar{G}_{21} + \bar{G}_{22} \tilde{Q}'(u) = -\frac{\rho}{2} (\tilde{T}_1 + \tilde{S}_1 \tilde{Q}'(u)).$$

Since

$$M(u) = D^{-1} \begin{pmatrix} u \\ \tilde{Q}(u) \end{pmatrix} \Rightarrow M'(u) = D^{-1} \begin{pmatrix} I_k \\ \tilde{Q}'(u) \end{pmatrix},$$

equations (57) reduce to

$$(58) \quad -\Gamma'(u) + G M'(u) = -\frac{\rho}{2} D^{-1} \begin{pmatrix} 0 \\ \tilde{T}_1 + \tilde{S}_1 \tilde{Q}'(u) \end{pmatrix},$$

where  $\tilde{T}_1 + \tilde{S}_1 \tilde{Q}'(u) = A_{22} \tilde{Q}'(u) - \tilde{Q}'(u) A_{11} + A_{21}$ . Since  $\tilde{Q}'(u) = D_{21} M'_1(u) + D_{22} \tilde{M}'(u)$  and  $M'_1(u) + D_{12} \tilde{M}'(u) = I_k$ , by using relations (10), we obtain that the

right side of (57) is equal to

$$-\frac{\rho}{2} \begin{pmatrix} H_1^{-1} \left( (I_k - H_1 M'_1(u)) A_{11} + \Lambda_1 H_1 M'_1(u) - \sum_{i=1}^m \lambda_i M'_i(u) \right) \\ H_2^{-1} \left( D_{12}^T A_{11} - H_2 \tilde{M}'(u) A_{11} + \Lambda_2 H_2 \tilde{M}'(u) - D_{12}^T \sum_{i=1}^m \lambda_i M'_i(u) \right) \end{pmatrix}.$$

Therefore, we get for every vector  $\Gamma_i(u) \in \mathbb{R}^k$ ,  $i = 1, \dots, m$ , that

(59)

$$-\Gamma'_i(u) + g_i M'_i(u) = -\frac{\rho}{2} \left( (h_i^{-1} - M'_i(u)) A_{11} + \lambda_i M'_i(u) - h_i^{-1} \sum_{i=1}^m \lambda_i M'_i(u) \right).$$

□

**Remark 3.2.** As previously observed in Remark 2.5, when  $A_{11} = 0$ , we have, for the second-order time derivatives,

$$\partial_{tt} u \sim t^{-2}, \quad \partial_{tt} \tilde{Z} \sim t^{-2}.$$

Therefore, in dealing with (37)-(38), we do not need to delete the second-order time derivatives  $u_{tt}$  and  $\tilde{Z}_{tt}$ , and relations (41)-(43) reduces to

$$(60) \quad C = 0, \quad \mathcal{C} = 0.$$

Besides, for each  $i = 1, \dots, m$  such that  $\lambda_i \neq 0$ , we can choose

$$(61) \quad g_i = \beta_1^i - \beta_{-1}^i = -\frac{1}{2\lambda_i} q_i, \quad \Gamma_i(u) = \mathcal{B}_1^i - \mathcal{B}_{-1}^i = g_i M_i(u).$$

Then, we can select  $q_i = |\lambda_i|$ , for  $i = 1, \dots, m$ , and we are free to choose

$$(62) \quad \beta_0^i = \frac{1}{2}, \quad \mathcal{B}_0^i(u) = \frac{M_i(u)}{2}.$$

Therefore, we obtain an upwind scheme for system (1), with the classical Roe upwinding approximation for the source term [21], namely

(63)

$$\begin{aligned} & \frac{f_{j,n+1}^i - f_{j,n}^i}{\Delta t} + \frac{\lambda_i}{2\Delta x} (f_{j+1,n}^i - f_{j-1,n}^i) - \frac{|\lambda_i|}{2\Delta x} \delta_x^2 f_{j,n}^i \\ &= \frac{M_i(u_{j-1}^n) + 2M_i(u_j^n) + M_i(u_{j+1}^n)}{4} + \frac{\text{sgn}(\lambda_i)}{4} (M_i(u_{j-1}^n) - M_i(u_{j+1}^n)) \\ & \quad - \frac{f_{j-1,n}^i + 2f_{j,n}^i + f_{j+1,n}^i}{4} - \frac{\text{sgn}(\lambda_i)}{4} (f_{j-1,n}^i - f_{j+1,n}^i), \end{aligned}$$

From now on, we shall refer to scheme (63) as the ROE scheme. Then, we have just proved the following result.

**Proposition 3.3.** *For  $A_{11} = 0$ , the local truncation error of the ROE scheme (63) verifies the time asymptotic estimate (44).*

#### 4. A MONOTONE TIME-AHO SCHEME FOR THE $2 \times 2$ CASE

In this section we apply the general result stated in Theorem 3.1 for the  $2 \times 2$  case described in the Example (2.1). For the artificial viscosity, we assume  $q_1 = q_2 = q$ ,  $q \in \mathbb{R}^+$ .

Hence, by applying the results given in Theorem 3.1, we get for the scheme parameters the following expressions:

$$(64) \quad C_1 = C_2 = -\frac{\rho}{2}, \quad \mathcal{C}_1(u) = -\frac{\rho}{2} M_1(u), \quad \mathcal{C}_2(u) = -\frac{\rho}{2} M_2(u)$$

$$(65) \quad \gamma_1 = \frac{q}{2a_+} + \frac{a\rho}{2a_+} (2\lambda + a), \quad \gamma_2 = -\frac{q}{2a_-} + \frac{a\rho}{2a_-} (2\lambda - a),$$

$$(66) \quad \Gamma_1(u) = \frac{q - a^2\rho}{2a_+} M_1(u) - \frac{a_-^2}{4\lambda} \rho u, \quad \Gamma_2(u) = -\frac{q - a^2\rho}{2a_-} M_2(u) + \frac{a_+^2}{4\lambda} \rho u.$$

To complete the definition of scheme (34) it is still necessary to choose four more free parameters, such as  $\mathcal{B}_0^{1,2}(\cdot)$  and  $\beta_0^{1,2}$ . For the  $2 \times 2$  case such parameters can be defined by applying the monotonicity conditions.

Starting from the scheme written in its diagonal form (32), the monotonicity conditions are given by the following relations, see [1]:

$$(67) \quad \begin{aligned} & \mathcal{B}'_{1;2,l}(\cdot) \geq 0 \quad \forall l = -1, 0, 1, \\ & 1 - \rho q + \Delta t \left( \mathcal{B}'_{1;2,0}(u) - \beta_0^{1,2} \right) \geq 0, \\ & \begin{cases} \frac{\rho\lambda_{1;2}}{2} + \frac{\rho q}{2} + \Delta t \left( \mathcal{B}'_{1;2,-1}(u) - \beta_{-1}^{1,2} \right) \geq 0, \\ -\frac{\rho\lambda_{1;2}}{2} + \frac{\rho q}{2} + \Delta t \left( \mathcal{B}'_{1;2,1}(u) - \beta_1^{1,2} \right) \geq 0. \end{cases} \end{aligned}$$

**Proposition 4.1** (Monotonicity). *Assume  $a > 0$  and  $\lambda \geq \|F'(u)\|_\infty$ . Under the assumptions of Theorem 3.1, the scheme (34) for the  $2 \times 2$  case verifies the monotonicity conditions (67) for the choices:*

$$(68) \quad \mathcal{B}_0^1(u) = M_1(u) - |\Gamma_1(u)| + \Delta x C^1(u), \quad \mathcal{B}_0^2(u) = M_2(u) - |\Gamma_2(u)| + \Delta x C^2(u).$$

$$(69) \quad \beta_0^1 = 1 - \gamma_1 + \Delta x c_1, \quad \beta_0^2 = 1 + \gamma_2 + \Delta x c_2,$$

under the CFL conditions

$$(70) \quad \Delta t \leq \min \left( \frac{1 - \lambda\rho}{1 + \gamma_1}, \frac{1 - \lambda\rho}{1 - \gamma_2} \right),$$

and, for  $\lambda > 2a$

$$(71) \quad \Delta x \leq \frac{a^2}{\lambda + a}, \quad \rho \leq \min \left( \frac{1}{\lambda}, \frac{2\lambda^2 m_-}{2a^2 \lambda m_- + a_+ a_-^2}, \frac{2\lambda^2 m_+}{2a^2 \lambda m_+ + a_-^2 a_+} \right),$$

where  $m_{1;2} = \min_u (M'_{1;2}(u)) > 0$ . Otherwise, if  $a < \lambda < 2a$ , we get the supplementary requirement

$$\rho \geq \frac{|\lambda - 2a|}{a^2 - \Delta x a_-}.$$

*Proof.* From consistency (33), we write

$$\mathcal{B}_{-1}^{1,2}(u) = \frac{1}{2} \left( M_{1;2}(u) - \mathcal{B}_0^{1,2}(u) - \Gamma_{1;2}(u) + \Delta x C_{1;2}(u) \right),$$

$$\mathcal{B}_1^{1,2}(u) = \frac{1}{2} \left( M_{1;2}(u) - \mathcal{B}_0^{1,2}(u) + \Gamma_{1;2}(u) + \Delta x C_{1;2}(u) \right),$$

$$\beta_{-1}^{1,2} = \frac{1}{2} \left( 1 - \beta_0^{1,2} - \gamma_{1;2} + \Delta x c_{1;2} \right), \quad \beta_1^{1,2} = \frac{1}{2} \left( 1 - \beta_0^{1,2} + \gamma_{1;2} + \Delta x c_{1;2} \right).$$

The first condition in (67) is equivalent to

$$(72) \quad M'_{1;2} - |\Gamma'_{1;2}| + \Delta x C'_{1;2} \geq \mathcal{B}'_{0,1;2} \geq 0,$$

and the third condition in (67) is equivalent to

$$(73) \quad \begin{cases} \frac{\rho}{2}(\pm\lambda + q) - \frac{\Delta t}{2}(1 - \beta_0^2 \mp \gamma_2 + \Delta x \, c_+) \geq 0, \\ \frac{\rho}{2}(\mp\lambda + q) - \frac{\Delta t}{2}(1 - \beta_0^1 \mp \gamma_1 + \Delta x \, c_-) \geq 0. \end{cases}$$

It is natural to assume that the CFL ratio  $\rho$  verifies the standard hyperbolic condition

$$(74) \quad \rho \leq \frac{1}{\lambda}.$$

Then, from relations (65), we have  $\gamma^+ \leq 0$  and  $\gamma^- \geq 0$  and for  $q \geq \lambda$ , we get monotonicity by choosing in (73),

$$(75) \quad \beta_0^1 = 1 - \gamma_1 + \Delta x \, c_-, \quad \beta_0^2 = 1 + \gamma_2 + \Delta x \, c_+,$$

under the limitation required in the second condition in (67),

$$(76) \quad \Delta t \leq \min \left( \frac{1 - \lambda\rho}{1 + \gamma_1}, \frac{1 - \lambda\rho}{1 - \gamma_2} \right).$$

To verify the request (72), we set first

$$(77) \quad \mathcal{B}'_{1,0}(u) = M'_1 - |\Gamma'_1| + \Delta x \, C'_1, \quad \mathcal{B}'_{2,0}(u) = M'_2 - |\Gamma'_2| + \Delta x \, C'_2.$$

For

$$\rho \leq \min \left( \frac{2\lambda^2 m_-}{2a^2 \lambda m_- + a_+ a_-^2}, \frac{2\lambda^2 m_+}{2a^2 \lambda m_+ + a_+^2 a_-} \right)$$

we get

$$\Gamma'_2 \leq 0, \quad \Gamma'_1 \geq 0,$$

then conditions (72) become

$$\left( 1 - \frac{\lambda}{2a_-} + \frac{\rho}{2} \left( \frac{a^2}{a_-} - \Delta x \right) \right) M'_2 + \frac{a_+^2}{4\lambda} \rho \geq 0,$$

$$\left( 1 - \frac{\lambda}{2a_+} + \frac{\rho}{2} \left( \frac{a^2}{a_+} - \Delta x \right) \right) M'_1 + \frac{a_-^2}{4\lambda} \rho \geq 0,$$

that are verified under the following limitations,

$$(78) \quad \lambda > \max(\|F'(u)\|_\infty, 2F'(0)), \quad \Delta x \leq \frac{a^2}{\lambda + a}.$$

If we choose

$$a < \lambda < 2a,$$

we get a limitation from the bottom for  $\rho$ ,

$$\rho \geq \frac{|\lambda - 2a|}{a^2 - \Delta x \, a_-}.$$

□

The monotone Time-AHO scheme has then the following expression,

$$\begin{aligned}
(79) \quad & \frac{f_{j,n+1}^1 - f_{j,n}^1}{\Delta t} - \frac{\lambda}{\Delta x} (f_{j+1,n}^1 - f_{j,n}^1) \\
&= \left(1 - \frac{\lambda - a^2 \rho}{2a_+}\right) M_1(u_{j,n}) + \frac{\lambda - a^2 \rho}{2a_+} M_1(u_{j+1,n}) - ((1 - \gamma^-) f_{j,n}^1 + \gamma^- f_{j+1,n}^1) \\
&\quad - \frac{\rho a^2}{4\lambda} (u_{j+1,n} - u_{j,n}) + \frac{\rho \Delta x}{2} (f_{j,n}^1 - M_1(u_{j,n})). \\
& \frac{f_{j,n+1}^2 - f_{j,n}^2}{\Delta t} + \frac{\lambda}{\Delta x} (f_{j,n}^2 - f_{j-1,n}^2) \\
&= \frac{\lambda - a^2 \rho}{2a_-} M_2(u_{j-1,n}) + \left(1 - \frac{\lambda - a^2 \rho}{2a_-}\right) M_2(u_{j,n}) - (|\gamma^+| f_{j-1,n}^2 + (1 - |\gamma^+|) f_{j,n}^2) \\
&\quad + \frac{\rho a^2}{4\lambda} (u_{j,n} - u_{j-1,n}) + \frac{\rho \Delta x}{2} (f_{j,n}^2 - M_2(u_{j,n})).
\end{aligned}$$

Notice that, the approximation of the source terms involve only the values of the solutions on the "upwinding-nodes", i.e.  $(x_j, x_{j+1})$  and  $(x_{j-1}, x_j)$  respectively for the first and the second equations.

The scheme may be then considered as an extension of the known approximation with the upwinding of the source term (63) described in Remark 3.2. Let us stress on the fact that, when  $F'(0) = 0$ , the local truncation error of the ROE scheme (63) verifies the decay properties obtained by the Time-AHO schemes in (44).

### 5. A MONOTONE TIME-AHO SCHEME FOR THE $3 \times 3$ CASE

In this section we compute a TAHO approximation for the  $3 \times 3$  case of example (2.2). In particular, we study the case where  $F(u) = a(u - u^2)$ , with  $a > 0$ . Then we have  $F'(0) = a > 0$ .

We recall that  $\alpha = |a| + \beta\lambda \in ]a, \lambda[$  and

$$(80) \quad \lambda - \alpha + |a| > |F'(u)|.$$

According to Proposition 3.1, we have  $A_{11} = a$ , and  $P = \lambda\alpha$ , and the coefficients  $h_i$  are

$$h_1 = \frac{2\lambda}{\alpha - a}, \quad h_2 = \frac{\lambda}{\lambda - \alpha}, \quad h_3 = \frac{2\lambda}{\alpha + a}.$$

We choose to take  $q_1 = q_3 = \lambda$ ,  $q_2 = 0$ . Consequently we find

$$(81) \quad g_1 = \frac{\lambda(1 - \rho\alpha)}{2(\lambda + a)} + \frac{\rho(\lambda + a)}{2}, \quad g_2 = -\frac{\rho(\lambda\alpha - a^2)}{2a}, \quad g_3 = -\frac{\lambda(1 - \rho\alpha)}{2(\lambda - a)} - \frac{\rho(\lambda - a)}{2}.$$

We have

$$(82) \quad \begin{cases} \Gamma_1(u) = \frac{\lambda(1 - \rho\alpha)}{2(\lambda + a)} M_1(u) + \frac{\rho(\alpha - a)}{4\lambda} (au - F(u)), \\ \Gamma_2(u) = -\frac{\rho\lambda\alpha}{2a} M_2(u) + \frac{\rho(\lambda - \alpha)}{2\lambda} (au - F(u)), \\ \Gamma_3(u) = -\frac{\lambda(1 - \rho\alpha)}{2(\lambda - a)} M_3(u) + \frac{\rho(\alpha + a)}{4\lambda} (au - F(u)). \end{cases}$$

For  $1 \leq i \leq 3$  we have

$$\beta_{\pm 1}^i = \frac{1}{2} \left(1 - \frac{\Delta t}{2} \pm g_i - \beta_0^i\right), \quad \mathcal{B}_{\pm 1}^i = \frac{1}{2} \left(M_i(u) \left(1 - \frac{\Delta t}{2}\right) \pm \Gamma_i(u) - \mathcal{B}_0^i(u)\right),$$

so it remains to determine the  $\beta_0^i$  and  $\mathcal{B}_0^i(u)$ . As for the  $2 \times 2$  case we use monotonicity criteria for those choices. From now on, we use the fact that  $a > 0$  and we consider only  $u \in [0, 1]$ , as this will be satisfied in our numerical test. We have then

$$(83) \quad -a \leq F'(u) \leq a,$$

$$(84) \quad 0 < 1 - \frac{\alpha}{\lambda} \leq M'_2(u) \leq 1 - \frac{\alpha - a}{\lambda} < 1,$$

$$(85) \quad 0 < \frac{\alpha - a}{2\lambda} \leq M'_i(u) \leq \frac{\alpha + a}{2\lambda} < 1, \quad i = 1, i = 3.$$

Hereafter we study the monotonicity conditions for each of the three equations obtained by applying the general numerical scheme (32) to this particular  $3 \times 3$  case. We denote

$$\mu = (\Delta x + 2a)(\lambda + a) - \lambda\alpha,$$

and

$$\nu_1 = \frac{\lambda}{2(\lambda - a)} \left( 1 - \frac{\alpha - a}{2\lambda} \right), \quad \nu_2 = -\frac{\lambda\alpha}{\lambda - a} \left( 1 - \frac{\alpha - a}{2\lambda} \right) + \frac{a(\alpha + a)}{\lambda} + \lambda - a.$$

**Proposition 5.1.** (*Monotonicity*). *Assume  $a > 0$ ,  $\lambda > 2a$ . We suppose that the following conditions are satisfied:*

$$(86) \quad \text{if } \nu_2 > 0 \quad \text{then} \quad \Delta x \leq \min \left( \frac{\lambda\alpha}{\lambda + a}, \frac{\lambda}{\nu_1 + \frac{\nu_2}{\alpha}} \right),$$

$$(87) \quad \text{if } \nu_2 \leq 0 \quad \text{then} \quad \Delta x \leq \min \left( \frac{\lambda\alpha}{\lambda + a}, \frac{\lambda}{\nu_1} \right),$$

$$(88) \quad \rho \leq \min \left( \frac{1}{\alpha}, \frac{1}{\lambda + \Delta x}, \frac{2a}{a\Delta x + \lambda\alpha}, \frac{2a(\lambda - \alpha)}{a(\lambda - \alpha)\Delta x + \alpha(\lambda^2 + \lambda - 2a^2)} \right),$$

$$(89) \quad \text{if } \mu > 0 \quad \text{then} \quad \rho \leq \frac{\lambda + 2a}{\mu},$$

$$(90) \quad \Delta t \leq \frac{1}{1 + \max_{u \in [0, 1]} |\Gamma'_2(u) - g_2|}.$$

Under the assumptions of Theorem 3.1, scheme (34) for the considered  $3 \times 3$  case is monotone for the choices:

$$(91) \quad \beta_0^1 = 1 - \frac{\Delta t}{2} - g_1, \quad \beta_0^2 = 1 - \frac{\Delta t}{2} + g_2, \quad \beta_0^3 = 1 - \frac{\Delta t}{2} + g_3,$$

$$(92) \quad \mathcal{B}_0^1(u) = M_1(u)(1 - \frac{\Delta t}{2}) - \Gamma_1(u), \quad \mathcal{B}_0^3(u) = M_3(u)(1 - \frac{\Delta t}{2}) + \Gamma_3(u),$$

and  $\mathcal{B}_0^2$  is a continuous function such that

$$(93) \quad -(1 - M'_2(u))(1 - \frac{\Delta t}{2}) - |\Gamma'_2(u) - g_2| = (\mathcal{B}_0^2)'(u) - \beta_0^2.$$

Such a function exists.

*Proof.* We do not detail the proof for equations 1 and 3, as it follows the lines of the  $2 \times 2$  case. Actually, under the above assumptions we have

$$g_1 > 0, \quad \Gamma'_1(u) \geq 0, \quad g_3 < 0, \quad \Gamma'_3(u) \leq 0.$$

Then we obtain a monotone TAHO scheme on those equations.



The treatment of the second equation is quite different. We first note that  $g_2 < 0$  but the sign of  $\Gamma'_2(u)$  does not depend on the parameters of the discretization. The monotonicity conditions are:

$$(94) \quad (\mathcal{B}_l^2)'(u) \geq 0, \quad l = -1, 0, 1,$$

$$(95) \quad -\beta_{\pm 1}^2 + (\mathcal{B}_{\pm 1}^2)'(u) \geq 0,$$

$$(96) \quad 1 - \Delta t(\beta_0^2 - (\mathcal{B}_0^2)'(u)) \geq 0.$$

The inequality (95) can be written as

$$(97) \quad -(1 - M'_2(u))(1 - \frac{\Delta t}{2}) - |\Gamma'_2(u) - g_2| \geq (\mathcal{B}_0^2)'(u) - \beta_0^2,$$

which is implied by equality (93). In the case under consideration, it is straightforward to determine the sign of  $\Gamma'_2(u) - g_2$  with respect to  $u$ . In our numerical experiment for example, we have  $u_0 \in ]0, 0.5[$  and  $u_1 \in ]0.5, 1[$  such that

$$\begin{aligned} \Gamma'_2(u) - g_2 &> 0 & \text{in} & [0, u_0[ \cup ]u_1, 1], \\ \Gamma'_2(u) - g_2 &< 0 & \text{in} & ]u_0, u_1[. \end{aligned}$$

We can therefore construct a continuous function  $\mathcal{B}_0^2$  such that

$$\begin{aligned} (\mathcal{B}_0^2)'(u) &= M'_2(u)(1 - \frac{\Delta t}{2}) - \Gamma'_2(u) + 2g_2 & \text{when} & \quad \Gamma'_2(u) - g_2 > 0, \\ (\mathcal{B}_0^2)'(u) &= M'_2(u)(1 - \frac{\Delta t}{2}) + \Gamma'_2(u) & \text{when} & \quad \Gamma'_2(u) - g_2 < 0. \end{aligned}$$

We set

$$\beta_{-1}^2 = -g_2, \quad \beta_0^2 = 1 - \frac{\Delta t}{2} + g_2, \quad \beta_1^2 = 0.$$

Therefore

$$-(1 - M'_2(u))(1 - \frac{\Delta t}{2}) - |\Gamma'_2(u) - g_2| = (\mathcal{B}_0^2)'(u) - \beta_0^2 < 0.$$

Consequently we have (95), (96) by (90), and (94) for  $l = \pm 1$ . It remains to satisfy

$$(98) \quad (\mathcal{B}_0^2)'(u) \geq 0.$$

**First case:**  $\Gamma'_2(u) - g_2 > 0$ .

By (83)-(84) we have

$$\begin{aligned} (\mathcal{B}_0^2)'(u) &\geq \frac{\lambda - \alpha}{\lambda} \left( 1 - \frac{\Delta t}{2} + \frac{\rho \lambda \alpha}{2a} \right) - \frac{\rho \alpha}{a \lambda} (\lambda^2 - a^2) \\ &\geq \frac{\lambda - \alpha}{\lambda} (1 - \sigma \rho). \end{aligned}$$

One can prove that  $\sigma > 0$ . We obtain (98) by (88).

**Second case:**  $\Gamma'_2(u) - g_2 < 0$ .

$$(\mathcal{B}_0^2)'(u) \geq M'_2(u) \left( 1 - \frac{\Delta t}{2} - \frac{\rho \lambda \alpha}{2a} \right) \geq 0$$

by (88). □

## 6. THE LINEAR CASE

**6.1. Numerical schemes.** Here we want to apply the argumentation of the above sections to the linear case, first considered in work [1]. We shall focus on problem (17) for  $F(u) = au$ . We set  $\alpha = \mu a$  and  $\beta = \mu$  and we study the following problem

$$(99) \quad \begin{cases} u_t + \alpha u_x + z_x = 0, \\ z_t + u_x - \alpha z_x = -\beta z. \end{cases}$$

The numerical approximation given in (34), here becomes

$$(100) \quad \begin{aligned} & \frac{U_j^{n+1} - U_j^n}{\Delta t} + \frac{A}{2\Delta x} (U_{j+1}^n - U_{j-1}^n) - \frac{Q}{2\Delta x} (U_{j+1}^n - 2U_j^n + U_{j-1}^n) \\ & = \mathcal{B}_{-1}U_{j-1}^n + \mathcal{B}_0U_j^n + \mathcal{B}_1U_{j+1}^n, \end{aligned}$$

where  $U = (u, z)^T$ , for  $\pm\xi = \pm\sqrt{1+\alpha^2}$  be the eigenvalues of matrix  $A$ ,  $Q = \text{diag}(\xi, \xi)$  and  $\mathcal{B}_{-1,0,1} = (\beta_{ij}^{-1,0,1})_{i,j=1,2}$  are the matrix of constant coefficients for the source term approximation.

First of all we shall analyze the decay properties of the local truncation error for the numerical approximations described in [1].

By Taylor expansion, the local truncation error for the numerical approximation (100) is given by

$$(101) \quad \left\{ \begin{aligned} \mathcal{T}_u &= \frac{\Delta t}{2}u_{tt} - \Delta x \left[ \frac{\xi}{2}u_{xx} + (\beta_{11}^1 - \beta_{11}^{-1})u_x + (\beta_{12}^1 - \beta_{12}^{-1})z_x + c_{11}u + c_{12}z \right] \\ &\quad + \mathcal{O}(\Delta x^2 + \Delta t^2), \\ \mathcal{T}_z &= \frac{\Delta t}{2}z_{tt} - \Delta x \left[ \frac{\xi}{2}z_{xx} + (\beta_{21}^1 - \beta_{21}^{-1})u_x + (\beta_{22}^1 - \beta_{22}^{-1})z_x + c_{21}u + c_{22}z \right] \\ &\quad + \mathcal{O}(\Delta x^2 + \Delta t^2), \end{aligned} \right.$$

where the  $(c_{ij})_{i,j=1,2}$  constants were defined in [1]. Let  $\Delta t/\Delta x = \rho$  be fixed. By relations

$$u_{tt} = \alpha^2 u_{xx} - (z_t - \alpha z_x)_x, \quad z_{tt} = \xi^2 z_{xx} - \beta(z_t + \alpha z_x),$$

we get, for some popular schemes, the following expansions.

(UP) for the source term point wise approximation, we have

$$(102) \quad \mathcal{B}_1 - \mathcal{B}_{-1} = 0, \quad C = 0,$$

$$(103) \quad \left\{ \begin{aligned} \mathcal{T}_u &= \frac{\Delta x}{2} [(\rho\alpha^2 - \xi)u_{xx} - \rho(z_t - \alpha z_x)_x] + \mathcal{O}(\Delta x^2) \sim \Delta x t^{-3/2}, \\ \mathcal{T}_z &= \frac{\Delta x}{2} [\xi(\rho\xi - 1)z_{xx} - \rho\beta(z_t + \alpha z_x)] + \mathcal{O}(\Delta x^2) \sim \Delta x t^{-3/2}. \end{aligned} \right.$$

(ROE) for the upwinding of the source term, we have

$$(104) \quad \mathcal{B}_1 - \mathcal{B}_{-1} = \frac{\beta}{2\xi} \begin{pmatrix} 0 & 1 \\ 0 & -\alpha \end{pmatrix}, \quad C = 0,$$

$$(105) \quad \left\{ \begin{aligned} \mathcal{T}_u &= \frac{\Delta x}{2} [(\rho\alpha^2 - \frac{(\xi^2-1)}{\xi})u_{xx} + (\frac{1}{\xi} - \rho)(z_t - \alpha z_x)_x] + \mathcal{O}(\Delta x^2) \sim \Delta x t^{-3/2}, \\ \mathcal{T}_z &= \frac{\Delta x}{2} [\xi(\rho\xi - 1)z_{xx} - \beta(\rho z_t + \alpha(\rho - \frac{1}{\xi})z_x)] + \mathcal{O}(\Delta x^2) \sim \Delta x t^{-3/2}. \end{aligned} \right.$$

(AHO2p) for the Asymptotic High Order scheme given in [1], we have

$$(106) \quad \mathcal{B}_1 - \mathcal{B}_{-1} = \frac{\beta\xi}{2} \begin{pmatrix} 0 & 1 \\ 1 & -2\alpha \end{pmatrix}, \quad C = \frac{\beta^2\xi}{2} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix},$$

$$(107) \quad \left\{ \begin{aligned} \mathcal{T}_u &= \frac{\Delta x}{2} [\rho\alpha^2 u_{xx} + (\xi - \rho)(z_t - \alpha z_x)_x] + \mathcal{O}(\Delta x^2) \sim \Delta x t^{-3/2}, \\ \mathcal{T}_z &= \frac{\Delta x}{2} [\xi(\rho\xi - 1)z_{xx} - \beta(\rho + \xi)(z_t + \alpha z_x)] + \mathcal{O}(\Delta x^2) \sim \Delta x t^{-3/2}. \end{aligned} \right.$$

Let us now go back to the Time-AHO schemes. According to the discussion of section 3.1, conditions given in propositions 3.1-4.1, here give rise to the following choice,

$$(108) \quad \mathcal{B}_1 - \mathcal{B}_{-1} = \frac{\beta\rho}{2} \begin{pmatrix} 0 & \xi - \alpha^2 \\ 1 & -2\alpha \end{pmatrix}, \quad C = \frac{\beta^2\rho}{2} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

For the local truncation error it then holds,

$$(109) \quad \begin{cases} \mathcal{T}_u = -\frac{\Delta x}{2} [\xi(1 - \rho\xi)(-z_{xt} + \alpha z_{xx})] + \mathcal{O}(\Delta x^2) \sim \Delta x t^{-2}, \\ \mathcal{T}_z = -\frac{\Delta x}{2} [\xi(1 - \rho\xi)z_{xx}] + \mathcal{O}(\Delta x^2) \sim \Delta x t^{-2}. \end{cases}$$

**Remark 6.1.** As for the non-linear case, for  $\alpha = 0$ , the Time-AHO scheme reduces to the Roe approximation, which in that case decays like in (109).

**6.2. Modified equation.** Here we wish to better understand the qualitative behavior of the numerical methods described in the previous section by applying the *modified equation method* (see for instance [18]). From Taylor expansion stated in (101), for all numerical schemes described in Section 6, we get the following modified system

$$(110) \quad \begin{cases} u_t + \alpha u_x + z_x = -\frac{\Delta t}{2} u_{tt} + \Delta x \left[ \frac{\xi}{2} u_{xx} + \gamma_{11} u_x + \gamma_{12} z_x + c_{11} u + c_{12} z \right], \\ z_t + u_x - \alpha z_x + \beta z = -\frac{\Delta t}{2} z_{tt} + \Delta x \left[ \frac{\xi}{2} z_{xx} + \gamma_{21} u_x + \gamma_{22} z_x + c_{21} u + c_{22} z \right]. \end{cases}$$

Since we are interested in long time simulations, we apply the decay rates results given in the analytical Section 2.2 to the derivatives terms of problem (110). As time increases, we can then take into account as modified equation the following *asymptotic modified system*,

$$(111) \quad \begin{cases} u_t + \alpha u_x + z_x = -\frac{\Delta t}{2} u_{tt} + \Delta x \left[ \frac{\xi}{2} u_{xx} + \gamma_{11} u_x + \gamma_{12} z_x + c_{11} u + c_{12} z \right], \\ u_x + \beta z = \Delta x [\gamma_{21} u_x + c_{21} u + c_{22} z]. \end{cases}$$

For all schemes described in Section 6, we get in the second equation  $[\gamma_{21} u_x + c_{21} u + c_{22} z] = 0$  and then, for all of them the asymptotic modified problem becomes

$$(112) \quad \begin{cases} u_t + \alpha u_x = \left( \frac{1}{\beta} + \Delta x \mathcal{D} \right) u_{xx} + O(\Delta x^2), \\ z = -\frac{1}{\beta} u_x + O(\Delta x^2), \end{cases}$$

where the constant  $\mathcal{D}$  depends on the selected scheme.

Specifically, for UP, ROE and AHO2p we obtain a perturbation of order  $O(\Delta x)$  on the diffusion term with

$$\mathcal{D}_{UP} = -\frac{1}{2} (\rho\alpha^2 - \xi), \quad \mathcal{D}_{ROE} = -\frac{1}{2} \left( \rho\alpha^2 - \frac{\xi^2 - 1}{\xi} \right), \quad \mathcal{D}_{AHO2p} = -\frac{\rho\alpha^2}{2};$$

while, for TAHO we get

$$\mathcal{D}_{TAHO} = 0.$$

The TAHO approximation is then second order accurate on the asymptotic diffusion (Chapman-Enskog) limit

$$(113) \quad \begin{cases} \hat{u}_t + \alpha \hat{u}_x = \frac{1}{\beta} \hat{u}_{xx}, \\ \hat{z} = -\frac{1}{\beta} \hat{u}_x. \end{cases}$$

## 7. NUMERICAL TESTS

In this section we show how, for large time simulations, Time-AHO schemes give better numerical results than standard approximations both for linear and non-linear cases.

From now on we shall call STD the following standard pointwise upwind approximation, for  $i = 1, \dots, m$

$$(114) \quad \frac{f_{j,n+1}^i - f_{j,n}^i}{\Delta t} + \frac{\lambda_i}{2\Delta x} (f_{j+1,n}^i - f_{j-1,n}^i) - \frac{|\lambda_i|}{2\Delta x} \delta_x^2 f_{j,n}^i = M_i(u_{j,n}) - f_{j,n}^i.$$

For all tests, we focus our attention on the numerical error as a function of time: denoting  $(u, Z)$  the conservative-dissipative variables, we plot the errors  $e_u(t) = \|(u^H - U^h)(t)\|_{L^\infty}$ ,  $e_z(t) = \|(Z^H - Z^h)(t)\|_{L^\infty}$  as the time  $t = n\Delta t$  increases, where  $(u^H, Z^H)$  is the reference solution obtained by the ROE scheme with  $\Delta x = \mathcal{O}(10^{-4})$ .

For all schemes, we fix the steps ratio  $\rho$  to verify all the CFL conditions; Since all schemes are of first order approximation, to emphasize the good behaviour of TAHO compared to the others schemes, we compute the numerical solutions  $U^h$  by using a quite big grid step  $\Delta x = \mathcal{O}(10^{-1})$ .

We then plot the different approximations of functions  $u$  and  $Z$  at final time  $T = 450$ , focusing on the point of maximum value of the solution to highlight the differences of the approximations. Near it, we show the most interesting plot of the  $l^\infty$  errors as a function of time.

Then, given different numerical approximations  $U^h$ , we look for constant  $C_u, \gamma_u, C_z, \gamma_z$  which best fit the equality

$$(115) \quad e_u(t) = \|(u^H - U^h)(t)\|_{L^\infty} = C_u t^{-\gamma_u}, \quad e_z(t) = \|(Z^H - Z^h)(t)\|_{L^\infty} = C_z t^{-\gamma_z}.$$

Given  $N$  data points  $(t_i, e(t_i))_{i=1,N}$ , we shall define  $\gamma$  and  $C$  as the solution of the following least squares problem,

$$\min_{C, \gamma} \sum_{i=1}^N |\ln(e(t_i)) - \ln(Ct^{-\gamma})|^2.$$

All numerical results we present show that for standard approximations, such as upwind (114) and ROE (63), the absolute error  $e(t)$ , for a fixed space step, decays as

$$e_u(t) = O(t^{-1/2}), \quad e_z(t) = O(t^{-1}),$$

while for the TAHO scheme, it improves of  $t^{-1/2}$  on the previous schemes.

**7.1. Results for the linear case.** Let us consider for system (99), the constant equilibrium state  $u = 1$  and  $z = 0$ . As in our previous work [1], we fix  $\beta = 5$  and we consider a small compactly supported perturbation of this constant solution as initial data.

$$(116) \quad u_0 = \chi_{[-1,1]}(-x^2 + 2), \quad z_0 = \chi_{[-1,1]}(-x^2 + 1).$$

We then compare the Time-AHO scheme (108) with the following schemes: the AHO2p scheme (106), the standard first-order point wise upwind scheme (102) and the ROE scheme (104).

**7.1.1. Test case with  $\alpha = 1$ .** As expected by our asymptotic analysis, the numerical results show a better performance of the TAHO scheme. In Figure 1-(a)-(b) we plot a zoom on the solutions  $u$  and  $z$  respectively, obtained by applying the different numerical schemes at final time  $T$ . The solution given by TAHO follows much better than the other the benchmark curve.

Moreover, always in Figure 1, the two plots (c) and (d) show the time evolution of the  $l^\infty$  errors  $e_u(t)$  and  $e_z(t)$  defined in (115) for all schemes considered; They

show how for the TAHO scheme, as time increases, both errors decay more quickly than the other. This result is also confirmed by Table 1, where the values of  $\gamma$  and  $C$  are computed. Looking at the different values of  $\gamma$ , it is clear that for the TAHO approximation the decay velocity of the absolute error improves of  $t^{-1/2}$  on the previous schemes.

To stress on the good behavior of the TAHO scheme, in Figure 2, we plot the solution  $u$  obtained by different numerical approximations with decreasing space step  $\Delta x$ . All schemes considered are of first order approximation, but looking at the numerical curves, it is clear how for large step the TAHO solution follows better the benchmark line than the other.

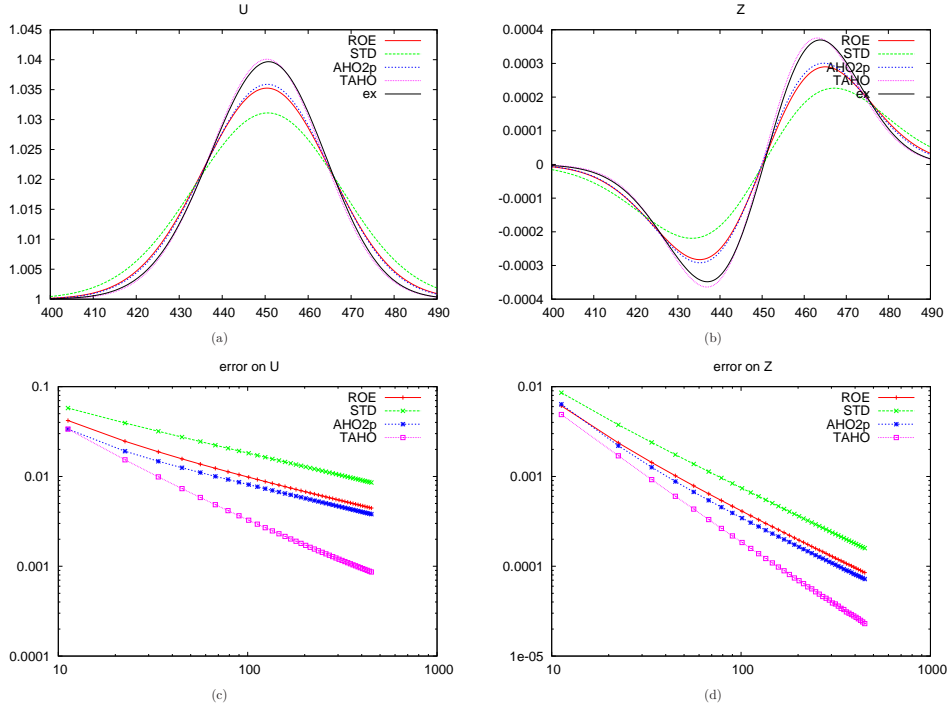


FIGURE 1. Linear Case Test, section 7.1.1. (a)-(b) zoom on the solutions  $u$  and  $z$  respectively obtained by the different schemes at final time  $T$ . The plot show a better performance of TAHO scheme. (c)-(d) time evolution of the  $l^\infty$  errors  $e_u(t)$  and  $e_z(t)$  defined in (115) for the different schemes. As expected by our asymptotic analysis, for the TAHO scheme the absolute errors  $e_{u,z}(t)$  decay faster as the time increases.

7.1.2. *Test case with  $\alpha = 0$ .* As previously observed in Remark 6.1, for  $\alpha = 0$  the ROE scheme (104) corresponds to our TAHO approximation.

For this particular case we can compare the TAHO/ROE scheme with the well-balanced approximation proposed by Gosse and Toscani in [10], which however is defined only in the case  $\alpha = 0$ . From now on we shall refer to this scheme as WB-GT.

Figure 3 shows the performances of TAHO/ROE, STD and WB-GT scheme for problem

$$(117) \quad \begin{cases} u_t + z_x = 0, \\ z_t + u_x = -\beta z, \end{cases}$$

| scheme | $C_u$    | $\gamma_u$ | $C_z$    | $\gamma_z$ |
|--------|----------|------------|----------|------------|
| STD    | 0.192993 | 0.510798   | 0.100412 | 1.058824   |
| ROE    | 0.143847 | 0.573922   | 0.073287 | 1.112843   |
| AHO2p  | 0.104897 | 0.547166   | 0.072874 | 1.143070   |
| TAHO   | 0.289768 | 0.961310   | 0.141281 | 1.432194   |

TABLE 1. Linear Case Test, section 7.1.1. Evaluation of constants  $\gamma$  and  $C$  for  $e_u(t) = C_u t^{-\gamma_u}$  and  $e_z(t) = C_z t^{-\gamma_z}$  defined in (115). For standard approximations, such as STD and ROE, the numerical results show that the absolute error decays as  $e_u(t) = O(t^{-1/2})$  and  $e_z(t) = O(t^{-1})$ ; while, for the TAHO scheme, it improves of  $t^{-1/2}$  on the previous schemes.

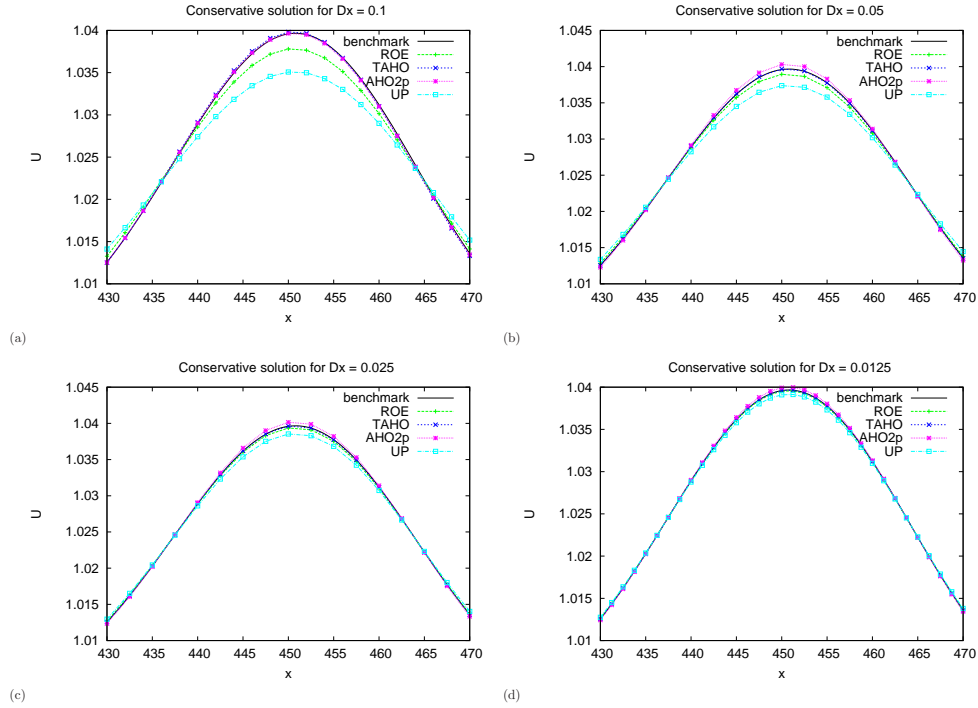


FIGURE 2. Linear Case Test, section 7.1.1. Zoom on the solutions  $U$  obtained by the different schemes at final time  $T$  by applying decreasing values of  $\Delta x$ : (a)  $\Delta x = 0.1$ , (b)  $\Delta x = 0.05$ , (c)  $\Delta x = 0.025$  and (d)  $\Delta x = 0.0125$ . The numerical solution obtained by TAHO scheme follows better than the others the benchmark curve already with a quite big space step.

with  $\beta = 5$ , at final time  $T = 450$ .

We observe that both TAHO/ROE and WB-GT give better performances than STD. The two numerical solutions obtained by TAHO/ROE and WB-GT showed respectively in Figure 3-(a)-(b) are overlapping and the numerical errors in Figure 3-(c)-(d) have a similar trend.

Indeed, according to the discussion of section 3.1, for the local truncation error of the WB-GT approximation it can be shown that

$$\mathcal{T}_u \sim \Delta x t^{-2}, \quad \mathcal{T}_z \sim \Delta x t^{-3/2}.$$

Then, we observe that the WB-GT scheme is Time-AHO with respect to the conservative variable  $u$ , while for the dissipative one it is not. The good behavior of WB-GT scheme may be confirmed by analyzing its order of convergence with respect the asymptotic modified problem (112). As done in section 6.2, when  $t$  goes to  $+\infty$ , it is possible to show that the WB-GT scheme is second order accurate with respect the asymptotic modified problem

$$(118) \quad \begin{cases} u_t = \frac{1}{\beta} u_{xx} + \frac{\beta \Delta x^2}{4(1 + \beta \Delta x/2)} u_{xx}, \\ z = -\frac{1}{\beta} u_x, \end{cases}$$

namely it is of second order with respect the Chapman-Enskog limit. Notice that, for  $\alpha \neq 0$ , it is possible to consider other Well Balanced scheme as in [7], but they are not THAO and their asymptotic performances are not of the same order (not shown).

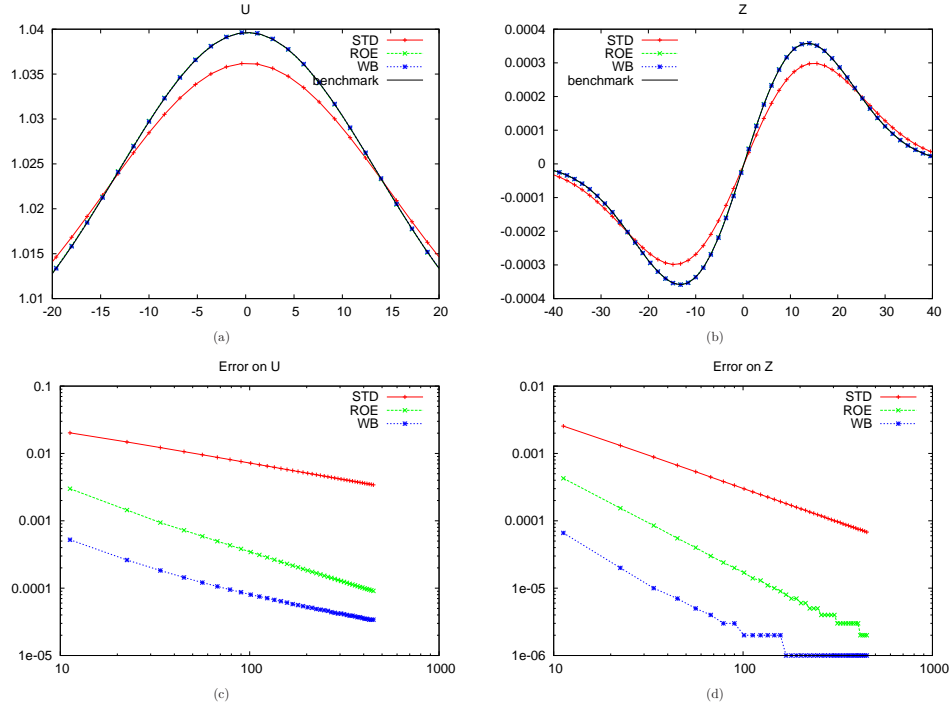


FIGURE 3. Linear Case Test for  $\alpha = 0$ , section 7.1.2. (a)-(b) zoom on the solutions  $u$  and  $z$  respectively obtained by the different schemes at final time  $T$  with  $\Delta x = 0.08$ . The plot show a better performance of TAHO/ROE and WB-GT scheme, with respect the STD approximation. As described in section 7.1.2 WB-GT scheme is Time-AHO with respect the conservative variables and, as TAHO/ROE, it is of order  $O(\Delta x^2)$  with respect the parabolic asymptotic state. (c)-(d) time evolution of the  $l^\infty$  errors  $e_u(t)$  and  $e_z(t)$  defined in (115) for the different schemes. As expected by our asymptotic analysis, for the TAHO/ROE and WB-GT schemes, the absolute errors  $e_{u,z}(t)$  decay faster as the time increases.

**7.2. Results for the non-linear  $2 \times 2$  test case.** We fix  $q = \lambda$  and we compare for the  $2 \times 2$  case the scheme TAHO (79) with ROE (63) and STD (114).

We shall consider the two different cases  $F'(0) \neq 0$  and  $F'(0) = 0$  and we select as initial datum the function

$$(119) \quad u_0 = \chi_{[-1,1]}(-x^2 + 1), \quad Z_0 = \frac{1}{\lambda} F(u_0(x)).$$

7.2.1. *The case  $F'(0) \neq 0$ .* Here we fix

$$F(u) = a(u - u^2)$$

and we compare our TAHO approximation (79) with STD and ROE schemes, defined in (114) and (63) respectively.

The numerical results show a better performance of the TAHO scheme. In Figure 4-(a)-(b) we plot a zoom on the solutions  $u$  and  $Z$  respectively, obtained by the different schemes at final time  $T$ . The solution given by applying the TAHO scheme follows much better than the other the benchmark curve. We stress on that the numerical solutions are computed with quite big step  $\Delta x = 0.1$ .

Moreover, always in Figure 4, the two plots (c) and (d) show the time evolution of the  $l^\infty$  errors  $e_u(t)$  and  $e_z(t)$  defined in (115) for all schemes considered; They show how for the TAHO scheme both errors decay as time increases more quickly than others. This result is also confirmed by Table 2, where the values of  $\gamma$  and  $C$  are computed. Looking at the different values of  $\gamma$ , it is clear that for the TAHO approximation the decay velocity of the absolute error improves of  $t^{-1/2}$  on the previous schemes.

| scheme | $C_u$    | $\gamma_u$ | $C_z$    | $\gamma_z$ |
|--------|----------|------------|----------|------------|
| STD    | 0.013797 | 0.374708   | 0.010744 | 0.341554   |
| ROE    | 0.004874 | 0.333634   | 0.007850 | 0.439996   |
| TAHO   | 0.111380 | 1.151517   | 0.495480 | 1.451030   |

TABLE 2. Non-Linear Case Test with  $F'(0) \neq 0$ , see section 7.2.1. Evaluation of constants  $\gamma$  and  $C$  for  $e_u(t) = C_u t^{-\gamma_u}$  and  $e_z(t) = C_z t^{-\gamma_z}$  defined in (115). For standard approximation STD and ROE, the absolute error decays as  $e_{u,z}(t) \approx O(t^{-1/2})$ ; while, for the TAHO scheme it improves of  $t^{-1/2}$ .

**7.3. Results for the  $3 \times 3$  system.** As initial data, we take the smooth function  $u_0$  defined by

$$u_0(x) = \chi_{[-1,1]} \exp \left( 1 - \frac{1}{1 - x^2} \right).$$

Then we set  $f_0(x) = M(u_0(x))$ . We know that in this case one has  $u(x, t) \in [0, 1]$  for all  $(x, t) \in \mathbb{R} \times [0, +\infty[$ . We choose  $a = 1$ ,  $\lambda = 2.1$ ,  $\beta = (\alpha - a)/\lambda = 0.1$ . The discretization parameters are  $\Delta x = 0.1$ ,  $\rho = \frac{1}{2\lambda}$ , which satisfy all the monotonicity requirements, see proposition 5.1.

The numerical results show a better performance of the TAHO scheme. In Figures 5, 6-(a)-(b), 7-(a)-(b), we plot the solutions  $u$ ,  $Z_1$  and  $Z_2$  respectively, obtained by the the STD, ROE and TAHO schemes at final time  $T$ , as well as the exact (reference) solution. The solution given by applying the TAHO scheme follows much better than the other the benchmark curve. We stress on that the numerical solutions are computed with quite big step  $\Delta x = 0.1$ .

Then in Figure 8-(a)-(b), we plot the time evolution of the  $l^\infty$  errors  $e_u(t)$  and  $e_z(t)$ . They show how for the TAHO scheme both errors decay as time increases more quickly than other. This result is also confirmed by Table 3, where the values of  $\gamma$  and  $C$  are computed. Looking at the different values of  $\gamma$ , it is clear that for



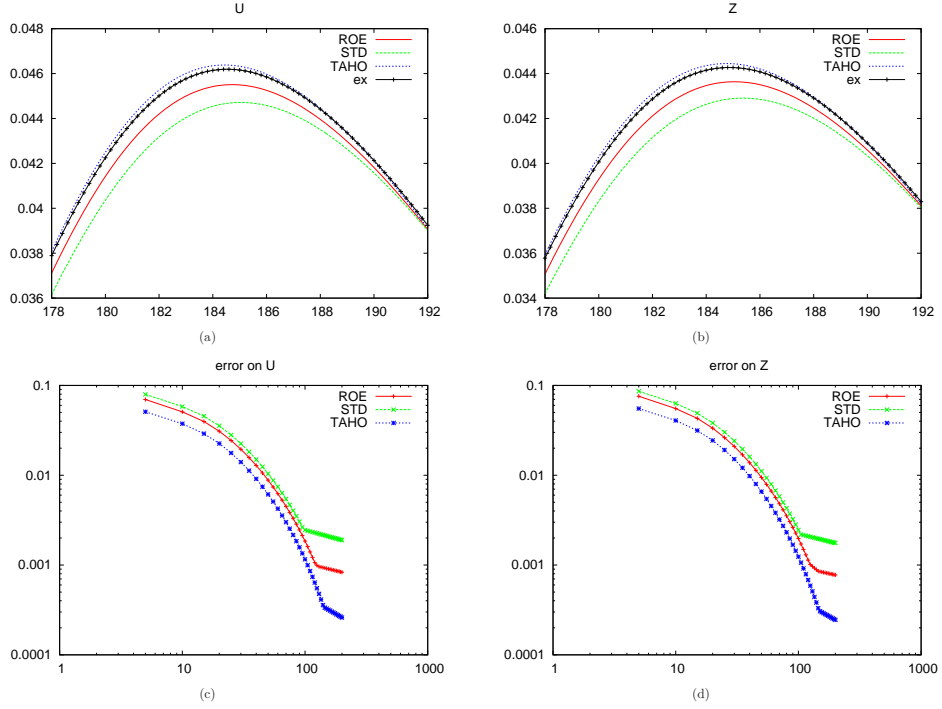


FIGURE 4. Non-Linear Case Test with  $F'(0) \neq 0$ , see section 7.2.1. (a)-(b) Zoom on the solutions  $u$  and  $Z$  respectively obtained by the different schemes at final time  $T$ . The plot show that TAHO scheme gives better results than others with a quite big step  $\Delta x = 0.1$ . (c)-(d) Time evolution of the  $l^\infty$  errors  $e_u(t)$  and  $e_z(t)$  defined in (115) for the different schemes. As expected by our asymptotic analysis, for the TAHO scheme the absolute errors  $e_{u,z}(t)$  decay faster as the time increases. This result is confirmed in Table 2, where we compute the decay parameters  $\gamma$  of absolute errors previously plotted.

the TAHO approximation the decay velocity of the absolute error improves of  $t^{-1/2}$  on the previous schemes.

| scheme | $C_u$  | $\gamma_u$ | $C_z$  | $\gamma_z$ |
|--------|--------|------------|--------|------------|
| STD    | 0.0052 | 0.54       | 0.0064 | 1.1        |
| ROE    | 0.0027 | 0.66       | 0.0036 | 1.2        |
| TAHO   | 0.006  | 1          | 0.012  | 1.62       |

TABLE 3. The  $3 \times 3$  system with  $F'(0) \neq 0$ , section 7.3. Evaluation of constants  $\gamma$  and  $C$  for  $e_u(t) = C_u t^{-\gamma_u}$  and  $e_z(t) = C_z t^{-\gamma_z}$  defined in (115). For STD and ROE approximations, the numerical results show that the absolute error decays as  $e_u(t) = O(t^{-1/2})$  and  $e_z(t) = O(t^{-1})$ ; while, for TAHO's it improves of  $t^{-1/2}$ .

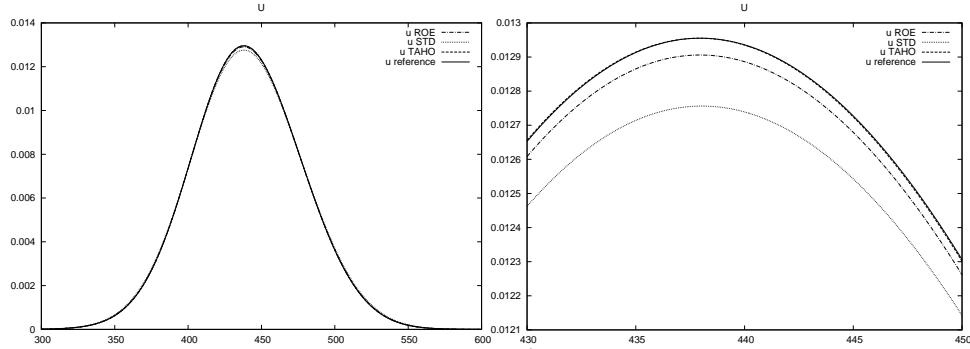


FIGURE 5. The  $3 \times 3$  system with  $F'(0) \neq 0$ . Left:  $u$  component at final time  $T$ . Right: detail. The reference and the TAHO computed solutions cannot be distinguished.

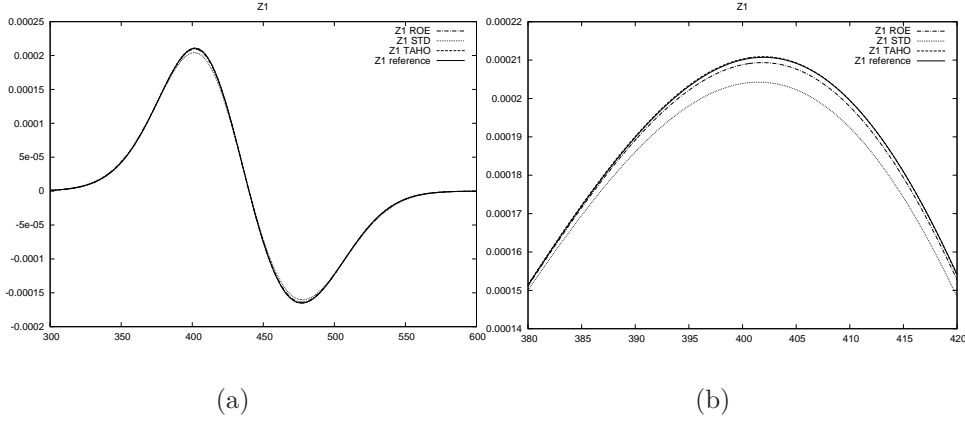


FIGURE 6. The  $3 \times 3$  system with  $F'(0) \neq 0$ . Left:  $Z_1$  component at final time  $T$ . Right: detail. The reference and the TAHO computed solutions cannot be distinguished.

## REFERENCES

- [1] D. AREGBA-DRIOLLET, M. BRIANI, R. NATALINI, *Asymptotic high-order schemes for  $2 \times 2$  dissipative hyperbolic systems*, SIAM J. Numer. Anal. **46** (2008), no.2, 869–894.
- [2] A. BERMUDEZ AND M.E. VASQUEZ, *Upwind methods for hyperbolic conservation laws with source terms*. Comput. and Fluids, 23 (1994), pp. 1049–1071.
- [3] S. BIANCHINI, B. HANOUEZ AND R. NATALINI, *Asymptotic behavior of smooth solutions for partially Dissipative hyperbolic systems with a convex entropy*, Communications Pure Appl. Math. **60** (2007), 1559–1622.
- [4] F. BOUCHUT, *Construction of BGK models with a family of kinetic entropies for a given system of conservation law*, J. Statist. Phys. 95 (1999), 113–170.
- [5] F. BOUCHUT, *Nonlinear stability of finite volume methods for hyperbolic conservation laws and well-balanced schemes for sources*. Birkhäuser, Frontiers in Mathematics series Birkhäuser, ISBN 3-7643-6665-6, 2004.
- [6] F. BOUCHUT, H. OUNAÏSSA, AND B. PERTHAME, *Upwinding of the source term at interfaces for Euler equations with high friction*, Comput. Math. Appl. 53 (2007), pp. 361–375.
- [7] L. GOSSE, *A well-balanced flux-vector splitting scheme designed for hyperbolic systems of conservation laws with source terms*, Comput. Math. Appl. 39 (2000), pp. 135–159.
- [8] L. GOSSE, *Localization effects and measure source terms in numerical schemes for balance laws*, Math. Comp. 71 (2002), 553–582.
- [9] L. GOSSE, F. JAMES, *Convergence results for an inhomogeneous system arising in various high frequency approximations*, Nmer. Math., 90 (2002), no 4, 721–753.

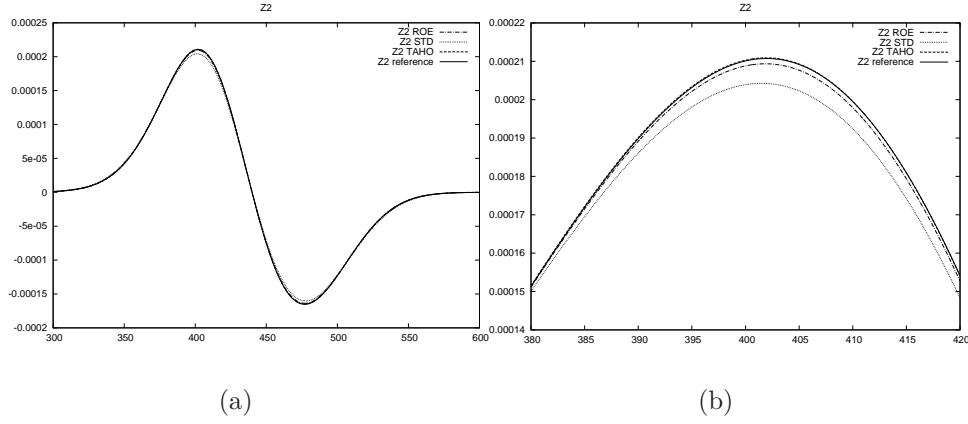


FIGURE 7. The  $3 \times 3$  system with  $F'(0) \neq 0$ . Left:  $Z_2$  component at final time  $T$ . Right: detail. The reference and the TAHO computed solutions cannot be distinguished.

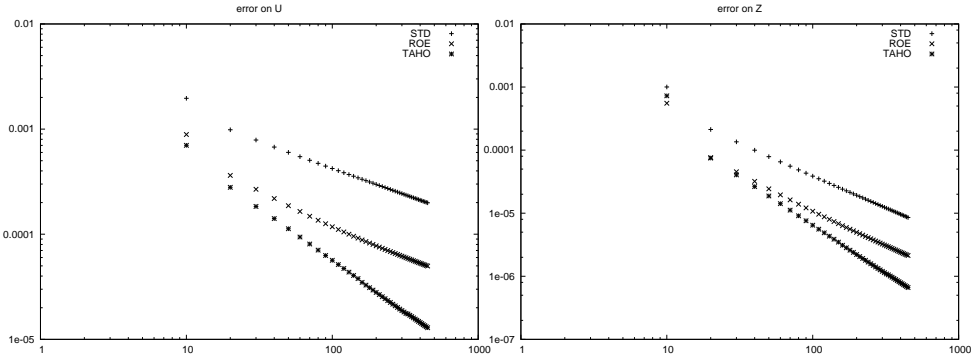


FIGURE 8. The  $3 \times 3$  system with  $F'(0) \neq 0$ . Absolute errors for  $u$  component (left) and  $Z$  components (right) with respect to time.

- [10] L. GOSSE, G. TOSCANI, *An asymptotic-preserving well-balanced scheme for the hyperbolic heat equations*, C. R. Math. Acad. Sci. Paris 334 (2002), no. 4, 337–342.
- [11] L. GOSSE, G. TOSCANI, *Space localization and well-balanced scheme for discrete kinetic models in diffusive regimes*, SIAM J. Numer. Anal. 41 (2003) 641–658.
- [12] J. M. GREENBERG AND A.-Y. LE ROUX, *A well-balanced scheme for the numerical processing of source terms in hyperbolic equations*, SIAM J. Numer. Anal. 33 (1996), pp. 1–16.
- [13] B. HANOUEZ, R. NATALINI, *Global existence of smooth solutions for partially dissipative hyperbolic systems with a convex entropy*, Arch. Ration. Mech. Anal. **169** (2003), no. 2, 89–117.
- [14] S. JIN, *Efficient asymptotic-preserving (AP) schemes for some multiscale kinetic equations*, SIAM J. Sci. Comput. 21 (1999), pp. 441–454.
- [15] S. JIN, *A steady-state capturing method for hyperbolic systems with geometrical source terms*, Math. Model. Num. Anal. 35 (2001), pp. 631–646.
- [16] JIN, SHI; SHI, YINGZHE, *A micro-macro decomposition-based asymptotic-preserving scheme for the multispecies Boltzmann equation*, SIAM J. Sci. Comput. 31 (2009/10), pp. 4580–4606.
- [17] JIN, SHI; XIN, ZHOU PING, *The relaxation schemes for systems of conservation laws in arbitrary space dimensions*, Comm. Pure Appl. Math. 48 (1995), no. 3, 235–276.
- [18] LEVEQUE, RANDALL J., *Numerical methods for conservation laws*. Lectures in Mathematics ETH Zürich, second edition (1992).
- [19] R. NATALINI, *Convergence to equilibrium for the relaxation approximations of conservation laws*, Comm. Pure Appl. Math. **49** (1996), no. 8, 795–823.

- [20] L. PARESCHI AND G. RUSSO, *Implicit-explicit Runge-Kutta schemes and applications to hyperbolic system with relaxation*, J. Sci. Comput. 25 (2005), pp. 129–155.
- [21] P. L. ROE, *Upwind differencing schemes for hyperbolic conservation laws with source term*, in Nonlinear Hyperbolic Problems, C. Carasso, P. A. Raviart, and D. Serre, eds., Lecture Notes in Math. 1270, Springer, Berlin, 1987, pp. 41–51.
- [22] Shizuta, Y.; Kawashima, S.; *Systems of equations of hyperbolic-parabolic type with applications to the discrete Boltzmann equation*, Hokkaido Math. J. 14 (1984) 435–457.